Essays on Minimal Supersolutions of BSDEs and on Cross Hedging in Incomplete Markets

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. Rer. Nat.)
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
der Humboldt-Universität zu Berlin

von
Herrn Dipl.-Math. Gregor Heyne

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Ulrich Horst
2. Prof. Dr. Peter Imkeller
3. Prof. Dr. Bruno Bouchard

Tag der Verteidigung: 30.08.2012

Abstract

In this thesis we study supersolutions of backward stochastic differential equations (BSDEs) and a specific hedging problem in mathematical finance.

In the first part of the thesis we analyze BSDEs with generators that are monotone in y, convex in z, jointly lower semicontinuous, and bounded below by an affine function of the control variable. We consider the set of all supersolutions with respect to a given generator and a given terminal condition. We prove several properties of this set such as stability under pasting and a certain downward directedness. The first central result establishes existence and uniqueness of a minimal supersolution. We show further that our setting allows to derive important properties of the minimal supersolution such as the flow property and the projectivity, with time consistency as a special case. Next we investigate the stability of the minimal supersolution with respect to pertubations of the generator or the terminal condition. We find that, for instance, the functional which maps the terminal condition to the infimum over all value processes evaluated at time zero is not only defined on the same domain as the original expectation operator, but also shares some of its main properties such as monotone convergence and Fatou's Lemma. Moreover, this leads to lower semincontinuity and dual representations of the functional. Finally, we demonstrate the scope of our method by giving a solution of the problem of maximizing expected exponential utility.

In the second part of the thesis we investigate quadratic hedging of contingent claims with basis risk.

We first show how to optimally cross-hedge risk when the logspread between the hedging instrument and the risk is asymptotically stationary. At the short end, the optimal hedge ratio is close to the cross-correlation of the log returns, whereas at the long end, the optimal hedge ratio equals one. For linear risk positions we derive explicit formulas for the hedge error, while for non-linear positions swift simulation analysis is possible. Finally, we demonstrate that even in cases with no clear-cut decision concerning the asymptotic stationarity of the logspread it is better to allow for mean reversion of the spread rather than to neglect it.

Secondly, we study a model where the correlation between the hedging instrument and the underlying of the contingent claim is random itself. We assume that the correlation is a process which evolves according to a stochastic differential equation with values between the boundaries -1 and 1. We keep the correlation dynamics general and derive an integrability condition on the correlation process and its first variation process that allows to describe and compute the quadratic hedge by means of a simple hedging formula. Furthermore we show that our conditions are fulfilled by a large class of correlation dynamics. We give various non-trivial explicit examples.

Zusammenfassung

In dieser Arbeit untersuchen wir Superlösungen von stochastischen Rückwärtsdifferentialgleichungen (BSDEs) und ein spezifisches Absicherungsproblem der Finanzmathematik.

Im ersten Teil der Arbeit analysieren wir BSDEs mit Generatoren, die monoton in y, convex in z, gemeinsam unterhalbstetig und von unten durch eine affine Funktion der Kontrollvariable beschränkt sind. Wir betrachten die Menge aller Superlösungen für einen fixen Generator und eine fixe Endbedingung. Wir beweisen mehrere Eigenschaften dieser Menge, wie zum Beispiel Stabilität bei Verkleben und eine bestimmte Gerichtetheit nach unten. Das erste Hauptresultat ist der Nachweis der Existenz und Eindeutigkeit einer minimalen Superlösung. Wir zeigen weiterhin, dass für die minimale Superlösung wichtige Eigenschaften, wie zum Beispiel die Flusseigenschaft und die Projektivität, mit Spezialfall Zeitkonsistenz, gelten. Danach untersuchen wir die Stabilität der minimalen Superlösung bezüglich Störungen des Generators oder der Endbedingung. Es stellt sich zum Beispiel heraus, dass das Funktional welches die Endbedingung auf das Infimum über alle Wertprozesse zur Zeit null abbildet nicht nur den gleichen Definitionsbereich wie der Erwartungswert hat, sondern auch einige seiner wichtigsten Eigenschaften, wie monotone Konvergenz und Fatou's Lemma teilt. Das führt im Weiteren zur Unterhalbstetigkeit und zu dualen Darstellungen dieses Funktionals. Schlussendlich demonstrieren wir die Bandbreite unserer Methode, indem wir eine Lösung des Nutzenmaximierungsproblems für die Exponentialnutzenfunktion herleiten.

Im zweiten Teil der Arbeit untersuchen wir die quadratische Absicherung von finanziellen Risikopositionen unter Basisrisiko.

Zuerst zeigen wir wie optimal abgesichert wird, wenn die Differenz der Logarithmen von Absicherungsinstrument und Risiko asymptotisch stationär ist. Am kurzen Ende ist die optimale hedge ratio nahe der Korrelation der logarithmierten Renditen, wohingegen am langen Ende die optimale hedge ratio gleich eins ist. Für lineare Risikopositionen leiten wir explizite Formeln für den Absicherungsfehler her und zeigen, dass für nichtlineare Positionen eine schnelle Simulation möglich ist. Schlussendlich, demonstrieren wir, dass es im Falle von Modellunsicherheit besser ist, mean reversion der logarithmischen Differenz von Absicherungsinstrument und Risiko anzunehmen.

Zweitens untersuchen wir ein Modell in dem die Korrelation zwischen Absicherungsinstrument und Basiswert stochastisch ist. Wir nehmen an, dass die Korrelation ein Prozess ist, der sich gemäß einer stochastischen Differentialgleichung mit Werten zwischen -1 und 1 entwickelt. Wir halten die Korrelationsdynamik allgemein und leiten eine Integrabilitätsbedingung bezüglich des Korrelationsprozesses und seines Prozesses erster Variation her, die uns erlaubt die optimale quadratische Absicherung durch eine einfache Formel zu beschreiben und zu berechnen. Weiterhin zeigen wir, dass unsere Bedingungen von einer großen Klasse von Korrelationsdynamiken erfüllt werden. Wir führen einige nichttriviale explizite Beispiele auf.

Contents

Introduction

In recent years the theory of backward stochastic differential equations (BSDEs) has emerged as an active field of research in probability theory. In mathematical terms the solution of a BSDE on a Brownian probability space is in principle a pair of adapted processes (Y, Z), which fulfills

$$Y_s - \int_s^t g_u(Y_u, Z_u)du + \int_s^t Z_u dW_u = Y_t, \qquad Y_T = \xi, \tag{I}$$

for all $0 \leq s \leq t \leq T$, where the $\mathcal{F}_T$-measurable random variable ξ is the so-called terminal condition and the measurable function g the generator. The process Y is usually referred to as the value process, while Z is the control process.

Ever since their inception the interest in these equations has been spurred not only by an intrinsic mathematical motivation but also because BSDEs appear naturally in a variety of problems in applied probability theory. There are, for instance, rich connections to partial differential equations, control theory and mathematical finance. Especially in the latter BSDEs are a powerful tool for providing solutions for example to questions on optimal hedging, utility maximization and stochastic equilibria. To illustrate the connection of BSDEs with mathematical finance think of the Bachelier model, see Bachelier [1900], with drift zero and volatility one, on a one-dimensional Brownian probability space. It is well-known that this model describes a complete financial market, where every contingent claim ξ is attainable. In the terminology of BSDEs we could describe the same fact by stating that the BSDE with generator $g = 0$ and terminal condition ξ has a solution (Y, Z). More precisely, Z yields the replicating strategy, whereas Y models the corresponding value process. This simple example of a zero generator additionally provides an insight into the relation of BSDEs and textbook stochastic analysis. Indeed, under the assumption that ξ is square integrable the unique solution of (I) is given by $(Y, Z) = (E[\xi \mid \mathcal{F}_\cdot], Z)$, where Z is obtained from the martingale representation theorem.

Let us give a brief and selective introduction to the development of BSDE theory. Stochastic backward differential equations with non-zero, linear generator made their first appearance in stochastic control theory as the equation satisfied by the adjoint process, see for example Bismut [1973]. The first systematic study was given in Pardoux and Peng [1990], who proved existence and uniqueness for Lipschitz continuous generators and square integrable terminal conditions. Following this work research on BSDEs increased considerably. One of the best known papers of this time is probably El Karoui et al. [1997b]. It presents several new results and collects existing knowledge into a comprehensive treatment. Moreover, the authors describe in great detail possible applications of Lipschitz BSDE theory in stochastic control theory, and in particular in mathematical finance. However, in many applications more sophisticated generators than Lipschitz continuous functions are required. Consequently, considerable work has been done to relax the assumptions on g and ξ, and to find conditions which guarantee existence and uniqueness of solutions, stability properties, and comparison theorems. With regards to this, a major milestone in BSDE theory is due to Kobylanski [2000]. In her work she gave existence, uniqueness and stability of BSDE with generators that have a quadratic growth in the control variable and a bounded terminal condition. Briand and Hu [2008] then proved that also unbounded terminal conditions with certain exponential integrability lead to unique solutions. These results opened the door for further applications, since problems arising in stochastic control often lead to BSDEs with generators with quadratic growth. For instance, preference based hedging in incomplete markets as considered in Hu et al. [2005] is a typical example. Another interesting but somewhat precluding result in this context is given in Delbaen et al. [2010]. For BSDEs with superquadratic growth in the control variable, it is essentially shown that, for bounded terminal conditions, in general a bounded solution does not need to exist and even if it exists it need not be unique.

In the first part of this thesis, we take the previous discussion, in particular the results in Delbaen et al. [2010], as a starting point. Instead of trying to find another set of conditions on the generator which allows for existence and uniqueness of BSDE solutions, we want to find a weaker concept than solutions of BSDEs which allows such theorems with less restrictions on the generator. We are especially interested in discontinuous generators with non-Lipschitz growth in the value variable, possibly superquadratic growth in the control variable and more general unbounded terminal conditions. It turns out that the notion of minimal supersolution of a BSDE is very well suited for this problem. In contrast to a solution of a BSDE a supersolution is a pair (Y, Z) of adapted process, which fulfills, for a given generator g and terminal condition ξ,

$$Y_s - \int_s^t g_u(Y_u, Z_u)du + \int_s^t Z_u dW_u \geq Y_t, \qquad Y_T \geq \xi, \tag{II}$$

for all $0 \leq s \leq t \leq T$. Moreover, the value process Y of a supersolution is required to be càdlàg. A supersolution is called minimal if its value process is, at any time, smaller than or equal to the value process of any other supersolution. The central questions regarding the set of supersolutions of a BSDE are as follows. Does there exist a minimal supersolution? Is it unique? How about stability of the minimal supersolution with respect to pertubations of the input data? Let us mention that supersolutions were introduced in El Karoui et al. [1997b], but no existence and uniqueness of the minimal supersolution was given. This was first done by Peng [1999], who, however, considered only Lipschitz continuous constrained generators and square integrable terminal conditions.

The main mathematical contribution of the first part of this thesis is to formulate a setting, which allows to extend the theory of supersolutions of BSDEs beyond Lipschitz continuous generators and to work with terminal conditions that are only integrable. More precisely, we consider generators that are convex with respect to z, monotone in y, jointly lower semicontinuous, and bounded below by an affine function of the control variable z. Given these assumptions we derive several new and important results. In particular we show that there exists a unique minimal supersolution and that it is stable with respect to pertubations of the terminal condition or the generator. For a more detailed and technical description of our novel approach, the results, and the precise mathematical contribution of this chapter we refer to page 5. Let us finally mention that the setting developed in the first part of this thesis is robust enough to allow even further progress in the theory of supersolutions. It is, for instance, possible to obtain existence and uniqueness results even if the convexity and the monotonicity assumptions are replaced by a mild normalization condition, see Heyne et al. [2012].

In the second part of this thesis we focus on a specific hedging problem in mathematical finance. Namely, we want to investigate how to hedge optimally if a hedging instrument is not perfectly correlated with the risk to be hedged, that is when a non-hedgeable risk, called basis risk remains. A typical example for such a situation is an airline company that wants to protect itself against changing kerosene prices. Since there is no liquid kerosene futures market the airline company may fall back on futures on less refined oil, such as crude oil futures, for hedging its kerosene risk. This is a reasonable approach, if the price evolvements of kerosene and of crude oil are very similar. Figure 0.1 illustrates the close comovement of the two price series at the IntercontinentalExchange (ICE). There is a rich literature about optimal hedging with

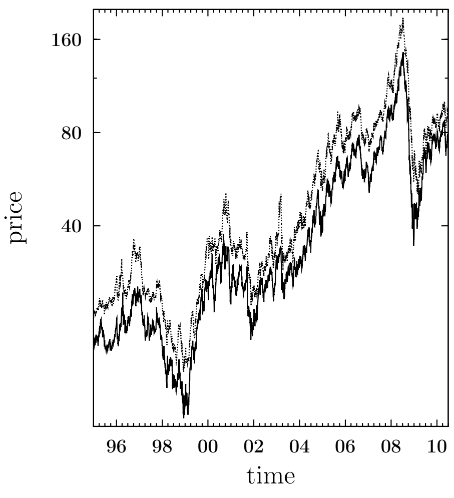

Figure 0.1.: Time evolvement of the daily price of crude oil in US$/BBL (dashed line) and for jet kerosene in US$/BBL (solid line) from 1995/01/02 until 2010/06/30 (resulting in 4043 observations).

basis risk, see for instance Duffie and Richardson [1991] and Schweizer [1992], who derive cross-hedging strategies minimizing quadratic objective functions, or Davis [2006] and Henderson [2002], who provide cross-hedging strategies maximizing the hedger's expected utility.

One of the goals, when studying cross hedging problems, is to derive strategies that allow for tractable representations. More precisely, in Markovian diffusion settings, hedging a European plain vanilla contingent claim, one often seeks to obtain characterizations of the optimal hedge ξ given by, at time t,

$$\xi_t = A_t^\top \nabla\psi(t, X_t). \tag{III}$$

Here X represents a vector of stochastic processes describing the financial model, $\psi(t,x)$ is a value function, typically an expectation of some functional of X_T, and A, often referred to as the hedge ratio, is a vector-valued stochastic process given by a function of the coefficients of the SDE describing X. Moreover, $\nabla\psi$ denotes the gradient of ψ with respect to the initial value of X and the components of this vector are commonly known as the Greeks. Now, for Formula (III) to be tractable it is essential that the Greeks can be computed easily. There exist a variety of methods, which allow for computation of $\nabla\psi(t,x)$, and often the particular choice of a method depends on the choice of model and vice versa. Popular approaches include, for instance, the finite difference method, the finite element method, the integration by parts method of Malliavin calculus, Fourier analysis based on affine model structure, and representations based on first variation processes.

The aim of the second part of this thesis is to motivate and to study two models that extend contemporary hedging literature in two different aspects. More precisely, we first investigate and interprete the effect of cointegration between risk and hedging instrument on the components of Formula (III) and the hedge error distribution. Secondly, we derive a tractable version of Formula (III) in a model where the correlation between risk and hedging instrument is random.

In Chapter 2 an empirical study shows that the correlation between the log prices in the kerosene and crude oil example above strongly depends on the sampling frequency. More precisely, the short-term correlations are considerably lower than the long-term correlations, pointing towards a long-term relationship with potential short-term deviations. However, this behaviour is not reflected by the widely studied models, where both processes, the risk source and the hedging instrument, are modeled as geometric Brownian motions. On the contrary, in these models, compare for example Duffie and Richardson [1991] and Schweizer [1992], and Davis [2006] and Henderson [2002], the variance of the spread of the log prices is increasing in time. Motivated by these observations we set up a new model for cross hedging, whose main feature is a mean reverting, or asymptotic stationary, spread between the log prices. In order to get a precise understanding of the influence an asymptotic stationary spread exerts on the hedging strategy we keep the model deliberately simple. By this the representation corresponding to Formula (III) can be calculated explicitly, which allows a rigorous study of the effect of a long-term relationship on optimal cross-hedging strategies. Moreover, our model allows an efficient calculation of the basis risk entailed by the optimal cross-hedges. A more detailed introduction, further motivation for the choice of our model, and the contributions of this chapter are given on page 8.

In Chapter 3 we change the focus from a model with asymptotic stationary spread to a model where both the risk source and the tradable asset are modeled as geometric Brownian motions. In all the related hedging literature, see above and also Musiela and Zariphopoulou [2004] and Monoyios [2004], the authors consider models where the correlation between the Brownian motions underlying both processes is assumed to be constant. The central aim of this chapter is to relax this restriction and to derive the existence of an optimal hedging strategy, when the correlation is allowed to be a random process with values between -1 and 1. More precisely, we will assume that the correlation process is the solution of a general stochastic differential equation (SDE), and we will work in the setting of local risk minimization, see Schweizer [2001] for an introduction. Given this framework, we want to find a representation of the locally risk minimizing strategy analogous to (III) and moreover explicitly characterize the corresponding Greeks. To this end, and here lie the first main mathematical contributions of this chapter, we prove differentiability of certain expectations, we prove explicit characterizations of the respective derivatives, and we prove an explicit representation for the control process of a particular BSDE. Our proofs require certain integrability assumptions on the correlation process and its first variation process. Further, in order to simplify the verification of these conditions for specific correlation models we prove sufficient and simple to check characterizations based directly on the coefficients of the SDE modeling the correlation. We find, in particular, that we may consider non-trivial correlation processes whose absolute value is not uniformly bounded away from one. A more detailed introduction to this chapter, its mathematical contribution and additional related literature are given on page 11.

The content of this thesis is strongly based on Drapeau et al. [2011], Ankirchner et al. [2011] and Ankirchner and Heyne [2012].

Let us give in the following more detailed descriptions of the results in this thesis.

Introduction to Chapter 1

On a filtered probability space, where the filtration is generated by a d-dimensional Brownian motion W, we consider the process $\hat{\mathcal{E}}^g(\xi)$ given by

$$\hat{\mathcal{E}}_t^g(\xi) = \operatorname{ess\,inf}\left\{Y_t \in L_t^0 : (Y,Z) \in \mathcal{A}(\xi,g)\right\}, \quad t \in [0,T],$$

where $\mathcal{A}(\xi,g)$ is the set of all pairs of càdlàg value processes Y and control processes Z, in some suitable space, such that

$$Y_s - \int_s^t g_u(Y_u,Z_u)du + \int_s^t Z_u dW_u \geq Y_t \quad \text{and} \quad Y_T \geq \xi, \tag{IV}$$

for all $0 \leq s \leq t \leq T$. Here the terminal condition ξ is a random variable, the generator g a measurable function of (y, z) and the pair (Y, Z) is a supersolution of the BSDE (IV).

The main objective of this chapter is to state a new and general set of assumptions on the generator and the terminal condition which guarantees that there exists a unique minimal supersolution. More precisely, we show that under our assumptions the process $\mathcal{E}^g(\xi) = \lim_{s\downarrow\cdot, s\in\mathbb{Q}} \hat{\mathcal{E}}^g_s(\xi)$ is well-defined, is a modification of $\hat{\mathcal{E}}^g(\xi)$, and equals the value process of the unique minimal supersolution, that is, there exists a unique control process $\hat{Z}$ such that $(\mathcal{E}^g(\xi), \hat{Z}) \in \mathcal{A}(\xi, g)$. Before we prove this central result we derive several properties of the set of supersolutions and the process $\hat{\mathcal{E}}^g(\xi)$, such as stability under pasting, downward directedness, and a comparison principle. Furthermore, we prove important properties such as the flow property and the projectivity, with time consistency as a special case. The existence theorem immediately yields a comparison theorem for minimal supersolutions. We also study the stability of the minimal supersolution with respect to pertubations of the terminal condition or the generator. Our results show that the mapping $\xi \mapsto \hat{\mathcal{E}}^g_0(\xi)$, which can be viewed as a nonlinear expectation, is not only defined on the same domain as the original expectation operator $E[\cdot]$, but also shares some of its main properties such as monotone convergence and Fatou's Lemma. These properties allow us to conclude that $\hat{\mathcal{E}}^g_0(\cdot)$ is L^1-lower semicontinuous and to obtain dual representation. Finally, we demonstrate the scope of our method by giving a solution of the problem of maximizing expected exponential utility.

Before we give further details on the specific choice of our spaces and assumptions, let us recall that related problems have been investigated throughout the literature before. Most notably, nonlinear expectations have been a prominent topic in mathematical economics since Allais famous paradox, see [Föllmer and Schied, 2004, Section 2.2]. Typical examples are the monetary risk measures introduced by Artzner et al. [1999] and Föllmer and Schied [2002], Peng's g-expectations, see Peng [1997], the variational preferences by Maccheroni et al. [2006], and the recursive utilities by Duffie and Epstein [1992]. Especially the g-expectation, which is defined as the initial value of the solution of a BSDE, is closely related to $\mathcal{E}^g_0(\cdot)$, since each pair (Y, Z) that solves the BSDE corresponding to (IV) is also a supersolution and hence an element of $\mathcal{A}(\xi, g)$. The concept of a supersolution of a BSDE appears already in El Karoui et al. [1997b, Section 2.2]. For further references see Peng [1999], who derives monotonic limit theorems for supersolutions of BSDEs and proves the existence of a minimal constrained supersolution. Another related concept are stochastic target problems, which were introduced and studied by Soner and Touzi [2002], by means of controlled stochastic differential equations and dynamic programming methods.

Our main contribution is to provide a setting where we relax the usual Lipschitz requirements for the generator g. As already mentioned, we suppose that g is convex with respect to z, monotone in y, jointly lower semicontinuous, and bounded below by an affine function of the control variable z. In order to provide an intuition on how these assumptions contribute toward the existence and uniqueness of a control process $\hat{Z}$ such that $(\mathcal{E}^g(\xi), \hat{Z}) \in \mathcal{A}(\xi, g)$, let us suppose for the moment that g is positive.

Given an adequate space of control processes, the value process of each supersolution and the process $\hat{\mathcal{E}}^g(\xi)$ are in fact supermartingales. By suitable pasting, we may now construct a decreasing sequence (Y^n) of supersolutions, whose pointwise limit is again a supermartingale and equal to $\hat{\mathcal{E}}^g(\xi)$ on all dyadic rationals. Since the generator g is positive, it can be shown that $\mathcal{E}^g(\xi)$ lies below $\hat{\mathcal{E}}^g(\xi)$, P-almost surely, at any time. This suggests to consider the càdlàg supermartingale $\mathcal{E}^g(\xi)$ as a candidate for the value process of the minimal supersolution. However, it is not clear a priori that the sequence (Y^n) converges to $\mathcal{E}^g(\xi)$ in some suitable sense. Yet, taking into account the additional supermartingale structure we can prove, by using Helly's theorem, that (Y^n) converges $P \otimes dt$-almost surely to $\mathcal{E}^g(\xi)$. It remains to obtain a unique control process $\hat{Z}$ such that $(\mathcal{E}^g(\xi), \hat{Z}) \in \mathcal{A}(\xi, g)$. To that end, we prove that, for monotone sequences of supersolutions, a positive generator yields, after suitable stopping, a uniform L^1-bound for the sequence of supremum processes of the associated sequence of stochastic integrals. This, along with a result by Delbaen and Schachermayer [1994], and standard compactness arguments and diagonalization techniques yield the candidate control process $\hat{Z}$ as the limit of a sequence of convex combinations. Now, joint lower semicontinuity of g, positivity, and convexity in z allow us to use Fatou's Lemma to verify that the candidate processes $(\mathcal{E}^g(\xi), \hat{Z})$ are a supersolution of the BSDE. Thus, $\mathcal{E}^g(\xi)$ is in fact the value process of the minimal supersolution and a modification of $\hat{\mathcal{E}}^g(\xi)$. Finally, the uniqueness of $\hat{Z}$ follows from the uniqueness of the Doob-Meyer decomposition of the càdlàg supermartingale $\mathcal{E}^g(\xi)$.

Let us give further reference of related assumptions and methods in the existing literature. Delbaen et al. [2010] consider superquadratic BSDEs with generators that are positive and convex in z but do not depend on y. However, their principal aim and their method differ from ours. Indeed, they primarily study the well-posedness of superquadratic BSDEs by establishing a dual link between cash additive time-consistent dynamic utility functions and supersolutions of BSDEs. To view supersolutions as supermartingales is one of the key ideas in our approach and we make ample use of the rich structure supermartingales provide. Note that the idea to use classical limit theory of supermartingales in the theory of BSDEs appears already in El Karoui and Quenez [1995], who study the problem of option pricing in incomplete financial markets. However, the analysis is done via dual formulations and only for linear generators that do not depend on y. The construction of solutions of BSDEs by monotone approximations is also a classical tool, see for example Kobylanski [2000] for quadratic generators and Briand and Hu [2008] for generators that are in addition convex in z. The idea to apply compactness theorems such as Delbaen and Schachermayer [1994, Lemma A1.1], or Delbaen and Schachermayer [1996, Theorem A], in order to derive existence of BSDEs is new to the best of our knowledge. Usually existence proofs rely on a priori estimates combined with a fixed point theorem, see for example El Karoui et al. [1997b], or on constructing Cauchy sequences in complete spaces, see for example Briand and Confortola [2008] or Ankirchner et al. [2007]. As already mentioned, Peng [1999] studies the existence and uniqueness of minimal supersolutions. However, he assumes a Lipschitz continuous constrained generator, a square integrable terminal condition, and employs a very different approach. It is based on a monotonic limit theorem, [Peng, 1999, Theorem 2.4], the penalization method introduced in El Karoui et al. [1997a], and it leads to monotone increasing sequences of supersolutions. Parallel to us, Cheridito and Stadje [2012] have investigated existence and stability of supersolutions of BSDEs. They consider generators that are convex in z and Lipschitz in y. However, their setting and methods are quite different from ours. Namely, they approximate by discrete time BSDEs and work with terminal conditions that are bounded lower semicontinuous functions of the Brownian motion. Finally, given our local L^1-bounds, the compactness underlying the construction of the candidate control process is a special case of results obtained by Delbaen and Schachermayer [1996].

Our second contribution is to allow for local supersolutions. This happens to be particularly adequate to establish monotone continuity properties of the minimal supersolution with respect to the terminal condition or the generator. We call a supersolution (Y, Z) of the BSDE local, if the stochastic integral of Z is well defined and thus a continuous local martingale. In order to avoid so-called "doubling strategies", present even for the most simple driver $g \equiv 0$, see Dudley [1977] or Harrison and Pliska [1981, Section 6.1], we require in addition that $\int Z dW$ is a supermartingale. In this setting, the stochastic integral of the candidate control process in the proof of the existence theorem is only a local martingale. However, we may once again use our assumptions on the generator to prove the supermartingale property. In addition, similar arguments allow us to formulate theorems such as montone convergence and Fatou's lemma for the non-linear operator $\hat{\mathcal{E}}_0^g(\cdot)$ on the same domain as the standard expectation $E[\cdot]$ and to obtain its L^1-lower semicontinuity. To complete the picture, we point out that our approach neither needs nor provides much integrability for the control processes. The underlying reason is that the compactness arguments in our proof are based on L^1 rather than $\mathcal{H}^1$-bounds for the stochastic integrals. Yet, given some additional integrability on the terminal condition, we obtain a candidate control process, whose stochastic integral belongs to $\mathcal{H}^1$ and therefore is a true martingale. However, monotone stability for an increasing sequence of terminal conditions does not hold without the additional assumption that $\mathcal{A}(\xi, g)$, where ξ is the limit terminal condition, is not empty. This guarantees the required integrability of the limit pair $(\hat{Y}, \hat{Z})$. In contrast, such an assumption is not necessary in our initial setting, where, in order to obtain suitable bounds and to construct the dominating candidate supermartingale, it is enough to know that the monotone limit of the minimal supersolutions at time zero is finite.

Replacing the positivity assumption with the condition that the generator is bounded below by an affine function of the control variable, it is obvious that the value and control processes of our supersolutions are supermartingales under another measure closely linked to the generator g. In fact, for a positive generator we have supermartingales with respect to the initial probability measure P, while for a non-positive generator, which is bounded below in the above sense, we consider supermartingales under the measure given by the corresponding Girsanov transform. This yields a generator dependent concept of admissibility. The implication thereof is illustrated by giving a minimal supersolution based approach to the well known problem of exponential expected utility maximization, where this admissibility is related in a natural way to the market price of risk.

The chapter is organized as follows. In Section 1.1 we fix our notations and the setting. A precise definition of minimal supersolutions, some of our main conditions and first structural properties of $\hat{\mathcal{E}}^g(\xi)$ are given in Section 1.2. Our main results, that is, existence and stability theorems, are given in Section 1.3, which concludes with an example on maximizing expected exponential utility.

Introduction to Chapter 2

The correlation between the price changes is the crucial determinant of an optimal cross-hedge. A common approach in the literature and in practice is to obtain the optimal hedge ratio by using the most frequent returns or price increments being available, irrespective of the time to maturity. This is a valid approach if the correlation (between the returns or price increments) and the ratio of the standard deviations are constant with respect to the sampling frequency, such as for correlated (geometric) Brownian motions. However, in many cases the correlation depends strongly on the selected time interval. For example in our empirical illustration the correlation of the daily log returns of kerosene and crude oil is only 0.52, which seems unexpectedly low given the strong comovement in the price series. The correlations of the weekly, monthly and yearly log returns in contrast are at 0.72, 0.84 and 0.98, respectively. Thus, the short-term correlation is considerably lower than the long-term correlation, pointing towards a long-term relationship with potential short-term deviations. This property is closely related to the concept of cointegration. It dates back to Engle and Granger [1987] and Granger [1981] and assumes that a set of time series share a long-term relationship with temporary deviations from this "equilibrium". More precisely, consider two integrated time series (of order one). They are cointegrated if a linear combination of them is stationary. This is supported for our example in Figure 0.2, which shows on the lower panel a clear mean reverting behavior of the spread between the logarithmic prices of kerosene and crude oil. Note that we do not use an estimated cointegration vector but rather assume that the spread between the log prices is stationary. This is more restrictive, but empirically supported by the p-value of the augmented Dickey-Fuller test (which is ≤ 0.001) indicating that the null hypothesis of a non-stationary spread is rejected. Kerosene and crude oil, however, is only one example for a pair of cointegrated processes and there are many studies pointing towards a cointegration relation between asset prices and corresponding hedging instruments, see e.g. Alexander [1999], Baillie [1989], Brenner and Kroner [1995], Lien and Luo [1993] and Ng and Pirrong [1996] and the references therein.

The long-term relationship between the kerosene price and the crude oil price leads to the observed increasing correlation in our example so that the optimal hedge ratios are not constant, but depend on time to maturity. Intuitively, for long-term hedges it is likely that the two assets are in their equilibrium relationship, whereas in the short-term the dynamics are dominated by noisy fluctuations.

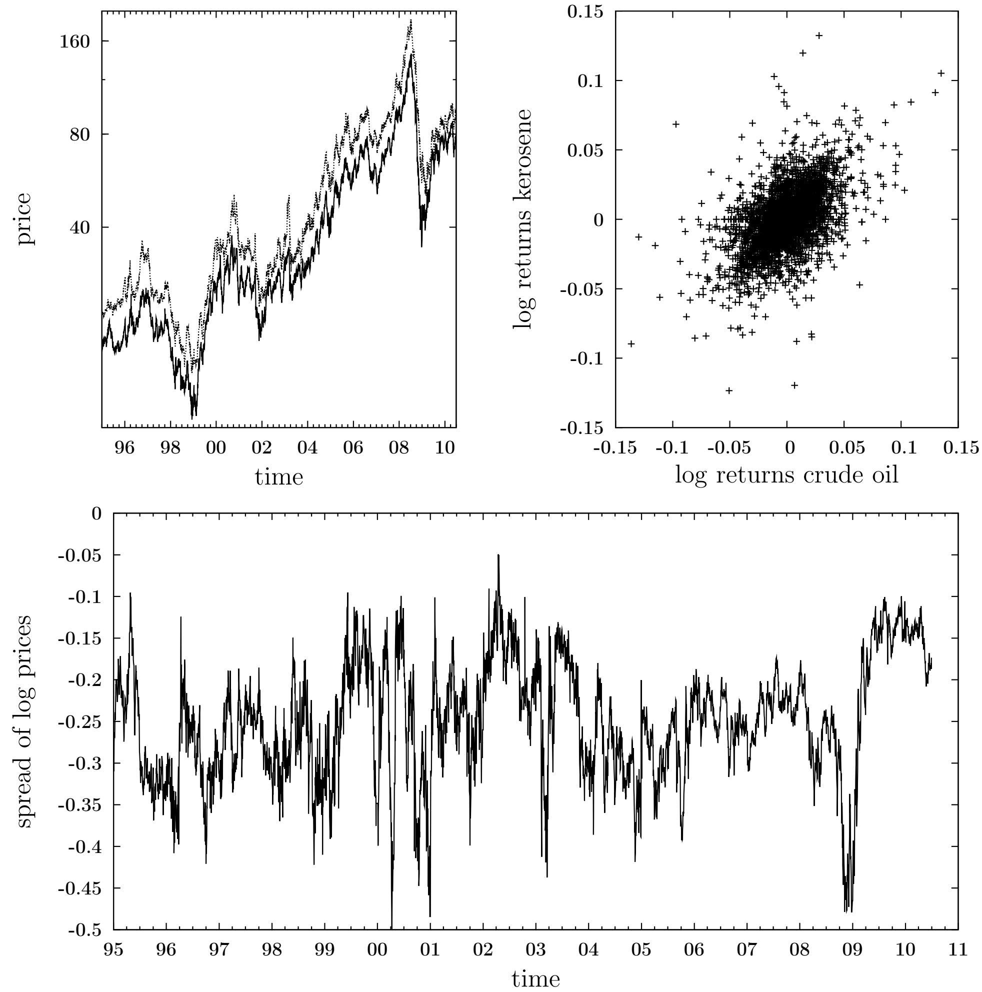

Figure 0.2.: The upper left panel depicts the time evolvement of the daily price of crude oil in US$/BBL (dashed line) and for jet kerosene in US$/BBL (solid line) from 1995/01/02 until 2010/06/30 (resulting in 4043 observations). The upper right panel exhibits the scatter plot of the corresponding daily log returns and shows that there is positive correlation among the two series as already mentioned in the text (with a correlation coefficient of 0.52). The lower panel depicts the time evolvement of the spread of the log prices.

The aim of this chapter is to set up a model that allows a rigorous study of the effect of a long-term relationship on optimal cross-hedging strategies, and at the same time allows an efficient calculation of the basis risk entailed by the optimal cross-hedges. We reproduce the long-term relationship of the prices by describing the logspread as an Ornstein-Uhlenbeck process, and by modeling the futures price as a geometric Brownian motion (GBM). Noteworthy, our model differs from the widely studied models where both processes, the risk source and the hedging instrument, are GBMs. Such models are considered for example in Duffie and Richardson [1991], Schweizer [1992], who derive cross-hedging strategies minimizing quadratic objective functions, and in Ankirchner et al. [2010], Davis [2006], Monoyios [2004], Musiela and Zariphopoulou [2004], who provide cross-hedging strategies maximizing the hedger's expected utility. In these models the spread of the log prices is not asymptotically stationary since the variance of the spread is proportional to time. We further show that these models underhedge the risk when cointegration is present (see Section 2.4). Our model, in contrast, explicitly accounts for an asymptotic stationary logspread. Furthermore, it is easy to estimate and it is still tractable enough to allow for a quick calculation of the hedge error standard deviation under different trading strategies. In particular, we are able to derive time-consistent strategies that minimize the variance of the hedge error.

To this end, we first solve the optimization problem of finding the dynamic strategy that minimizes the variance of the hedge error. Variance optimal hedging strategies have been first described in Föllmer and Sondermann [1986]. We make use of their method and transfer it to the specific case of cross-hedging risk with futures contracts within our Markovian model. The optimal hedging strategy can be expressed in terms of the risk's Greeks and a hedge ratio decaying with time to maturity. Moreover, for linear risk positions we are able to derive a closed-form formula for the hedge error standard deviation.

The chapter is structured as follows. Section 2.1 introduces our model and presents some empirical evidence, while Section 2.2 briefly reviews hedging with futures contracts and derives the variance optimal hedging strategy for our model. Section 2.3 develops the implied hedge errors within our model for linear and non-linear positions and Section 2.4 compares the hedge errors between different models and (suboptimal) hedging strategies emphasizing the importance of allowing for an asymptotic stationary spread. An extension of our model to account for stochastic volatility is given in Section 2.5. Section 2.6 concludes.

Introduction to Chapter 3

In this chapter we assume that the price of the tradable asset and the value of the non-tradable index evolve according to geometric Brownian motions. However, we will assume that the correlation between the driving Brownian motions is not constant, but a random process with values between -1 and 1. More precisely, we will assume that the correlation process is the solution of a stochastic differential equation (SDE).

We consider European options on the non-tradable index and derive the asset hedging strategy that locally minimizes the quadratic hedging error, the so-called locally risk minimizing strategy. Essentially, the optimal hedge can be described by the following factors: the asset hedge ratio, defined as

$$\rho_t \frac{\text{index vola}}{\text{asset vola}},$$

where ρ_t is the correlation process, and the correlation hedge ratio, defined as

$$\gamma \frac{\text{correlation vola}}{\text{asset vola}},$$

where γ is the correlation between the asset and ρ_t. The derivative with respect to the asset (resp. the correlation) of the expected value of the option under the so-called minimal equivalent local martingale measure will be called asset delta (resp. correlation delta). We will show that the optimal hedge is the asset hedge ratio multiplied with the asset delta plus the correlation hedge ratio multiplied with the correlation delta, that is

$$\begin{aligned} \text{optimal hedge} = \text{asset hedge ratio} \times \text{asset delta} & \\ + \ \text{correlation hedge ratio} \times \text{correlation delta}. & \end{aligned}$$

In order to obtain this characterization we first show that in our setting the locally risk minimizing strategy may be expressed in terms of the solution (Y, Z) of a certain linear BSDE. In fact, this is an observation made before in El Karoui et al. [1997b]. More precisely, the locally risk minimizing strategy depends on the control process Z of the corresponding BSDE. Now, in order to obtain an explicit representation of the strategy one has to explicitly characterize Z. In El Karoui et al. [1997b] it is shown that in Markovian settings, that is, when the randomness in the terminal condition and in the generator of a BSDE is induced by a forward diffusion, in principle such representations are possible. More precisely, it can be shown that under suitable regularity assumptions on the coefficients of the forward diffusion process, the terminal condition, and the generator the control process of a BSDE can be expressed as a function of the first variation process of the value process Y and the volatility coefficient of the forward process. However, the coefficients of the volatility matrix of the forward processes in our model do not satisfy the prerequisites of El Karoui et al. [1997b, Proposition 5.9]. In particular, our coefficients do not have uniformly bounded derivatives, and therefore these results are not directly applicable in our setting.

With this in mind there are two main mathematical contributions in this chapter. Firstly, we prove that the value process of our linear BSDE can be differentiated with respect to the initial values of our forward processes and that the gradient can be explicitly written as a vector of expectations based on first variation processes of the forward diffusions. In general, by assuming a stochastic correlation, there is no closed formula for the asset delta, but it is straightforward to show that it has a representation in terms of a simple expectation based on first variation processes. The major effort, however, lies in showing that the correlation delta can be expressed as a simple expectation as well. Here the main difficulty is that the dynamics of the non-tradable asset contains a term which may induce the non-differentiability of the flow of the non-tradable asset with respect to the initial value of the correlation process and which has to be controlled when interchanging expectation and differentiation. In order to fix this, we require that the absolute value of the correlation process is always strictly below one and we introduce a natural integrability condition on the correlation and its first variation process. Given these conditions various technical arguments yield that expectation and differentiation may be interchanged. Secondly, we recover in our setting the classical representation of Z, despite the lack of regularity of our coefficients. This requires to prove a characterization of the control process of our particular BSDEs in the spirit of El Karoui et al. [1997b, Proposition 5.9]. To that end we rely on a technical proof which is based on an approximation and mollification procedure, and on arguments involving Malliavin calculus. Both steps combined yield the explicit characterization of the optimal strategy.

Given the existence and the representation of the locally risk minimizing strategy a natural question is then which correlation processes fulfill our main assumptions? Based on boundary theory for diffusions and integrability properties of solutions of linear SDEs we prove that our conditions can be sufficiently characterized by conditions based directly on the coefficients of the SDE modeling the correlation. We use this characterization to provide several examples of correlation processes which fulfill our assumptions. In particular, we may consider non-trivial correlation processes whose absolute value is not uniformly bounded away from one.

We want to point out to two papers that allow for stochastic correlation in pricing contingent claims. In van Emmerich [2006] quanto options are priced by assuming that the exchange rate is stochastically correlated with the underlying. Frei and Schweizer [2008] deal with exponential utility indifference valuation of contingents claims based on risk sources that are stochastically correlated with assets traded on financial markets. However, both only give results on the value of the optimization problem but do not have explicit expressions for the optimal hedging strategy, compare, for instance, the remark at the end of Section 3.1 in Frei and Schweizer [2008].

The chapter is organised as follows. In Section 3.1 we give a short introdution into local risk minimization. Section 3.2 introduces our model and gives an overview on the main results we obtained. The details we use to derive our hedge formula are provided in Section 3.3. We continue in Section 3.4 by analyzing the boundary behaviour and integrability properties of correlation processes. We conclude with Section 3.5 by giving some explicit examples of correlation processes for which our main results hold.

Acknowledgments

I thank my supervisor, Ulrich Horst, for his guidance, support and continuous encouragement during my time as a PhD student. Special thanks go to Stefan Ankirchner for the fruitful collaboration on cross hedging and for many intense discussions. I also owe special thanks to Michael Kupper for sharing his research enthusiasm and creativity with me. Moreover, I am very grateful to Samuel Drapeau for his relentless pursuit of an optimal presentation of stochastic target problems. I thank Georgi Dimitroff and Christian Pigorsch for sharing their interest on cross hedging with stationary spread.

I also thank Markus Mocha, Anthony Réveillac, Mikhail Urusov, and Nicholas Westray for many helpful discussions.

Parts of my research were done while being employed at the Humboldt Universität zu Berlin and the Quantitative Products Laboratory, a joint initiative by Deutsche Bank AG, Humboldt Universität zu Berlin and Technische Universität Berlin. I gratefully acknowledge their financial support which enabled me to concentrate on my research.

Part I.

Minimal Supersolutions of Convex BSDEs

1. Minimal Supersolutions of Convex BSDEs

On a filtered probability space, where the filtration is generated by a d-dimensional Brownian motion W, we consider the process $\hat{\mathcal{E}}^g(\xi)$ given by

$$\hat{\mathcal{E}}^g_t(\xi) := \operatorname*{ess\,inf}\left\{Y_t \in L^0_t : (Y,Z) \in \mathcal{A}(\xi,g)\right\}, \quad t \in [0,T],$$

where $\mathcal{A}(\xi,g)$ is the set of all pairs of càdlàg *value processes* Y and *control processes* Z such that

$$Y_s - \int_s^t g_u(Y_u,Z_u)du + \int_s^t Z_u dW_u \geq Y_t \quad \text{and} \quad Y_T \geq \xi, \tag{1.1}$$

for all $0 \leq s \leq t \leq T$. Here the *terminal condition* ξ is a random variable, the *generator* g a measurable function of (y,z) and the pair (Y,Z) is a *supersolution* of the *backward stochastic differential equation (BSDE)* (1.1).

The main objective of this chapter is to state conditions which guarantee that there exists a unique minimal supersolution. More precisely, we show that the process $\mathcal{E}^g(\xi) = \lim_{s\downarrow\cdot, s\in\mathbb{Q}} \hat{\mathcal{E}}^g_s(\xi)$ is a modification of $\hat{\mathcal{E}}^g(\xi)$ and equals the value process of the unique minimal supersolution, that is, there exists a unique control process $\hat{Z}$ such that $(\mathcal{E}^g(\xi),\hat{Z}) \in \mathcal{A}(\xi,g)$. The existence theorem immediately yields a comparison theorem for minimal supersolutions. We also study the stability of the minimal supersolution with respect to pertubations of the terminal condition or the generator. Our results show that the mapping $\xi \mapsto \hat{\mathcal{E}}^g_0(\xi)$, which can be viewed as a nonlinear expectation, fulfills a montone convergence theorem and Fatou's Lemma on the same domain as the expectation operator $E[\cdot]$. These properties allow us to conclude that $\hat{\mathcal{E}}^g_0(\cdot)$ is L^1-lower semicontinuous.

The chapter is organized as follows. In Section 1.1 we fix our notations and the setting. A precise definition of minimal supersolutions, some of our main conditions and first structural properties of $\hat{\mathcal{E}}^g(\xi)$ are given in Section 1.2. Our main results, that is, existence and stability theorems, are given in Section 1.3. In Section 1.3.4 we illustrate the scope of our method with an example on maximizing expected exponential utility. We conclude this chapter with a version of Helly's theorem in Section 1.4.

1.1. Setting and Notations

We consider a fixed time horizon $T > 0$ and a filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in[0,T]},P)$, where the filtration $(\mathcal{F}_t)$ is generated by a d-dimensional Brownian motion W and fulfills the usual conditions. We further assume that $\mathcal{F} = \mathcal{F}_T$. The set of $\mathcal{F}$-measurable and $\mathcal{F}_t$-measurable random variables is denoted by L^0 and L^0_t, respectively, where random variables are identified in the P-almost sure sense. The sets L^p and L^p_t denote the set of random variables in L^0 and L^0_t, respectively, with finite p-norm, for $p \in [1, +\infty]$. Throughout this chapter, inequalities and strict inequalities between any two random variables or processes X^1, X^2 are understood in the P-almost sure or in the $P \otimes dt$-almost sure sense, respectively, that is, $X^1 \leq (<) X^2$ is equivalent to $P\left[X^1 \leq (<) X^2\right] = 1$ or $P \otimes dt\left[X^1 \leq (<) X^2\right] = 1$, respectively. Given a process X and $t \in [0,T]$ we denote $X^*_t := \sup_{s\in[0,t]} |X_s|$. We denote by $\mathcal{T}$ the set of stopping times with values in $[0,T]$ and hereby call an increasing sequence of stopping times (τ^n), such that $P[\bigcup_n \{\tau^n = T\}] = 1$, a *localising sequence of stopping times*. By $\mathcal{S} := \mathcal{S}(\mathbb{R})$ we denote the set of all càdlàg progressively measurable processes Y with values in $\mathbb{R}$ and further denote with $Prog$ the σ-algebra on $\Omega \times [0,T]$ generated by all progessively measurable processes. For $p \in [1, +\infty[$, we further denote by $\mathcal{L}^p := \mathcal{L}^p(W)$ the set of progressively measurable processes Z with values in $\mathbb{R}^{1\times d}$, such that $\|Z\|_{\mathcal{L}^p} := E[(\int_0^T Z_s^2 ds)^{p/2}]^{1/p} < +\infty$. For any $Z \in \mathcal{L}^p$, the stochastic integral $(\int_0^t Z_s dW_s)_{t\in[0,T]}$ is well defined, see [Protter, 2004], and is by means of the Burkholder-Davis-Gundy inequality a continuous martingale. For the $\mathcal{L}^p$-norm, the set $\mathcal{L}^p$ is a Banach space, see [Protter, 2004]. We further denote by $\mathcal{L} := \mathcal{L}(W)$ the set of progressively measurable processes with values in $\mathbb{R}^{1\times d}$, such that there exists a localising sequence of stopping times (τ^n) with $Z1_{[0,\tau^n]} \in \mathcal{L}^1$, for all $n \in \mathbb{N}$. Here again, the stochastic integral $\int ZdW$ is well defined and is a continuous local martingale.

For adequate integrands a, Z, we generically write $\int ads$ or $\int ZdW$ for the respective integral processes $(\int_0^t a_s ds)_{t\in[0,T]}$ and $(\int_0^t Z_s dW_s)_{t\in[0,T]}$. Finally, given a sequence (x_n) in some convex set, we say that a sequence (y_n) is in the *asymptotic convex hull* of (x_n), if $y_n \in conv\{x_n, x_{n+1}, \ldots\}$, for all n.

Normal integrands have been introduced by Rockafellar [1976] and are particularly adequate to model integral problems with constraints. In our setting, a normal integrand is a function $g : \Omega \times [0,T] \times \mathbb{R} \times \mathbb{R}^{1\times d} \to]-\infty, +\infty]$, such that

- $(y,z) \mapsto g(\omega,t,y,z)$ is lower semicontinuous for all $(\omega,t) \in \Omega \times [0,T]$;
- $(\omega,t) \mapsto g(\omega,t,y,z)$ is progressively measurable for all $(y,z) \in \mathbb{R} \times \mathbb{R}^{1\times d}$.

It is shown in [Rockafellar and Wets, 1998, Chapter 14.F], that for all progressively measurable processes Y, Z, the process $g(Y,Z)$ is itself progressively measurable and so, the integral $\int g(Y,Z)\,ds$ is well defined P-almost surely under the assumption that $+\infty - \infty = +\infty$. The section theorem as well as the Fubini, Tonelli theorem [Kallenberg, 2002, Lemma 1.26 and Theorem 1.27] extend to that context. Finally, the lower semicontinuity yields an extended Fatou's lemma, that is,

$$\int \liminf_n g_s(Y^n_s, Z^n_s)\,ds \leq \liminf_n \int g_s(Y^n_s, Z^n_s)\,ds,$$

for any sequence Y^n, Z^n of progressively measurable processes, if $g \geq 0$.

1.2. Minimal Supersolutions of BSDEs

1.2.1. Definitions

Given a normal integrand g, henceforth called *generator*, and a *terminal condition* $\xi \in L^0$, a pair $(Y,Z) \in \mathcal{S} \times \mathcal{L}$ is a *supersolution* of a BSDE, if, for all $s,t \in [0,T]$, with $s \leq t$, holds

$$Y_s - \int_s^t g_u(Y_u, Z_u)du + \int_s^t Z_u dW_u \geq Y_t \quad \text{and} \quad Y_T \geq \xi. \tag{1.2}$$

For such a supersolution (Y,Z), we call Y the *value process* and Z its *control process.* Due to the càdlàg property, Relation (1.2) holds for all stopping times $0 \leq \sigma \leq \tau \leq T$, in place of s and t, respectively. Note that the formulation in (1.2) is equivalent to the existence of a càdlàg increasing process K, with $K_0 = 0$, such that

$$Y_t = \xi + \int_t^T g_u(Y_u, Z_u)du + (K_T - K_t) - \int_t^T Z_u dW_u, \qquad t \in [0,T]. \tag{1.3}$$

Although the notation in (1.3) is standard in the literature concering supersolutions of BSDEs, see for example El Karoui et al. [1997b], Peng [1999], we will keep with (1.2) since the proofs of our main results exploit this structure. We consider only those supersolutions $(Y,Z) \in \mathcal{S} \times \mathcal{L}$ of a BSDE where Z is *admissible*, that is, where the continuous local martingale $\int ZdW$ is a supermartingale. We are then interested in the set

$$\mathcal{A}(\xi, g) = \{(Y,Z) \in \mathcal{S} \times \mathcal{L} : Z \text{ is admissible and (1.2) holds}\} \tag{1.4}$$

and the process

$$\hat{\mathcal{E}}^g_t(\xi) = \operatorname{ess\,inf}\left\{Y_t \in L^0_t : (Y,Z) \in \mathcal{A}(\xi, g)\right\}, \quad t \in [0,T]\,. \tag{1.5}$$

By $\hat{\mathcal{E}}^g$ we mean the functional mapping terminal conditions $\xi \in L^0$ to the process $\hat{\mathcal{E}}^g(\xi)$. Since the set $\mathcal{A}(\xi, g)$ and therefore $\hat{\mathcal{E}}^g(\xi)$ depends on the time horizon T, we indicate this by writing $\mathcal{A}_T(\xi, g)$ and $\hat{\mathcal{E}}^g_{\cdot,T}(\xi, g)$, if necessary. Note that the essential infima in (1.5) can be taken over those $(Y,Z) \in \mathcal{A}(\xi, g)$, where $Y_T = \xi$. A pair (Y,Z) is called *minimal supersolution*, if $(Y,Z) \in \mathcal{A}(\xi, g)$, and if for any other supersolution $(Y', Z') \in \mathcal{A}(\xi, g)$, holds $Y_t \leq Y'_t$, for all $t \in [0,T]$.

1.2.2. General Properties of $\mathcal{A}(\cdot, g)$ and $\hat{\mathcal{E}}^g$

In this section we collect various statements regarding the properties of $\mathcal{A}(\cdot, g)$ and $\hat{\mathcal{E}}^g$. The first lemma ensures that the set of admissible control processes is stable under pasting and that we may concatenate elements of $\mathcal{A}(\xi, g)$ along stopping times and partitions of our probability space.

Lemma 1.1. *Fix a generator g, a terminal condition $\xi \in L^0$, a stopping time $\sigma \in \mathcal{T}$, and $(B^n) \subset \mathcal{F}_\sigma$ a partition of Ω.*

1. *Let $(Z^n) \subset \mathcal{L}$ be admissible. Then $\bar{Z} = Z^1 1_{[0,\sigma]} + \sum_{n\geq 1} Z^n 1_{B_n} 1_{]\sigma,T]}$ is admissible.*
2. *Let $((Y^n, Z^n)) \subset \mathcal{A}(\xi, g)$ such that $Y^1_\sigma 1_{B^n} \geq Y^n_\sigma 1_{B^n}$, for all $n \in \mathbb{N}$. Then $(\bar{Y}, \bar{Z}) \in \mathcal{A}(\xi, g)$, where*

$$\bar{Y} = Y^1 1_{[0,\sigma[} + \sum_{n\geq 1} Y^n 1_{B^n} 1_{[\sigma,T]} \quad \text{and} \quad \bar{Z} = Z^1 1_{[0,\sigma]} + \sum_{n\geq 1} Z^n 1_{B^n} 1_{]\sigma,T]}. \tag{1.6}$$

Proof. 1. Let M^n and $\bar{M}$ denote the stochastic integrals of the Z^n and $\bar{Z}$, respectively. It follows from $(Z^n) \subset \mathcal{L}$ and from (B_n) being a partition that $\bar{Z} \in \mathcal{L}$ and that $\int_{s\vee\sigma}^{t\vee\sigma} \bar{Z}_u dW_u = \sum 1_{B_n} \int_{s\vee\sigma}^{t\vee\sigma} Z^n_u dW_u$. Now observe that the admissibility of all Z^n yields

$$E\left[\bar{M}_t - \bar{M}_s \mid \mathcal{F}_s\right] = E\left[M^1_{(t\wedge\sigma)\vee s} - M^1_s \mid \mathcal{F}_s\right] + E\left[\sum_{n\geq 1} 1_{B_n} E\left[M^n_{t\vee\sigma} - M^n_{s\vee\sigma} \mid \mathcal{F}_{s\vee\sigma}\right] \mid \mathcal{F}_s\right] \leq 0\,,$$

for $0 \leq s \leq t \leq T$.

2. $\bar{Z}$ is admissible by Item 1. Since $Y^1_\sigma 1_{B^n} \geq Y^n_\sigma 1_{B^n}$, for all $n \in \mathbb{N}$, it follows on the set $\{s < \sigma \leq t\}$ that

$$\begin{aligned}
Y^1_s - \int_s^\sigma g_u(Y^1_u, Z^1_u)du + \int_s^\sigma Z^1_u dW_u - \int_\sigma^t g_u(\bar{Y}_u, \bar{Z}_u)du + \int_\sigma^t \bar{Z}_u dW_u& \\
\geq Y^1_\sigma - \sum_{n\geq 1} 1_{B^n}\left(\int_\sigma^t g_u(Y^n_u, Z^n_u)du - \int_\sigma^t Z^n_u dW_u\right)& \\
\geq \sum_{n\geq 1} 1_{B^n}\left(Y^n_\sigma - \int_\sigma^t g_u(Y^n_u, Z^n_u)du + \int_\sigma^t Z^n_u dW_u\right) \geq \sum_{n\geq 1} 1_{B^n} Y^n_t&.
\end{aligned}$$

Hence,

$$\bar{Y}_s - \int_s^t g_u(\bar{Y}_u, \bar{Z}_u)du + \int_s^t \bar{Z}_u dW_u \geq 1_{\{\sigma > t\}} Y^1_t + \sum_{n\geq 1} 1_{B^n}\left(1_{\{\sigma\leq s\}} Y^n_t + 1_{\{s<\sigma\leq t\}} Y^n_t\right) = \bar{Y}_t$$

and thus $(\bar{Y}, \bar{Z}) \in \mathcal{A}(\xi, g)$. □

In the following, some properties of the generator are key, and therefore, we say that a generator g is

(Pos) positive, if $g(y, z) \geq 0$, for all $(y, z) \in \mathbb{R} \times \mathbb{R}^{1\times d}$.

(Inc) increasing, if $g(y,z) \geq g(y',z)$, for all $y, y' \in \mathbb{R}$ with $y \geq y'$, and all $z \in \mathbb{R}^{1\times d}$.

(Dec) decreasing, if $g(y,z) \leq g(y',z)$, for all $y, y' \in \mathbb{R}$ with $y \geq y'$, and all $z \in \mathbb{R}^{1\times d}$.

The next proposition addresses how $\mathcal{A}(\xi, g)$ depends on the terminal condition and the generator and which impact they have on $\hat{\mathcal{E}}^g(\xi)$. The first two properties are crucial in the proof of the existence and uniqueness theorem in Section 1.3. The third item concerns the monotonicity of $\hat{\mathcal{E}}^g(\xi)$ with respect to ξ and g. Combined with the existence theorem, this yields in fact a comparison principle for minimal supersolutions of BSDEs. We will illustrate its scope in the proof of our stability results and in the example on utility maximization in Section 1.3.4. Finally, the last item concerns the cash (super/sub) additivity of the functional $\hat{\mathcal{E}}^g(\xi)$.

Proposition 1.2. *For $t \in [0,T]$, generators g, g' and terminal conditions $\xi, \xi' \in L^0$, it holds*

1. *the set $\{Y_t : (Y,Z) \in \mathcal{A}(\xi, g)\}$ is directed downwards.*
2. *assuming* (Pos), *$\xi^- \in L^1$ and $\mathcal{A}(\xi, g) \neq \emptyset$, then for all $\varepsilon > 0$, there exists $(Y^\varepsilon, Z^\varepsilon) \in \mathcal{A}(\xi, g)$ such that $\hat{\mathcal{E}}^g_t(\xi) \geq Y^\varepsilon_t - \varepsilon$.*
3. (monotonicity) *if $\xi' \leq \xi$ and $g'(y,z) \leq g(y,z)$, for all $y, z \in \mathbb{R} \times \mathbb{R}^{1\times d}$, then $\mathcal{A}(\xi', g') \supset \mathcal{A}(\xi, g)$ and $\hat{\mathcal{E}}^{g'}_t(\xi') \leq \hat{\mathcal{E}}^g_t(\xi)$.*
4. (convexity) *if $(y,z) \mapsto g(y,z)$ is jointly convex, then $\mathcal{A}(\lambda\xi + (1-\lambda)\xi', g) \supset \lambda\mathcal{A}(\xi, g) + (1-\lambda)\mathcal{A}(\xi', g)$, for all $\lambda \in (0,1)$, and so*
$$\hat{\mathcal{E}}^g_t(\lambda\xi + (1-\lambda)\xi') \leq \lambda\hat{\mathcal{E}}^g_t(\xi) + (1-\lambda)\hat{\mathcal{E}}^g_t(\xi').$$
5. *for $m \in L^0_t$,*
   - (cash superadditivity) *assuming* (Inc) *and $m \geq 0$, then $\hat{\mathcal{E}}^g_t(\xi + m) \geq \hat{\mathcal{E}}^g_t(\xi) + m$.*
   - (cash subadditivity) *assuming* (Dec)*, $m \geq 0$, and the existence of $(Y,Z) \in \mathcal{A}(\xi, g)$, such that $\mathcal{A}_t(Y_t + m, g) \neq \emptyset$, then $\hat{\mathcal{E}}^g_t(\xi + m) \leq \hat{\mathcal{E}}^g_t(\xi) + m$.*
   - (cash additivity) *assuming that g does not depend on y, the existence of $(Y,Z) \in \mathcal{A}(\xi, g)$, such that $\mathcal{A}_t(Y_t + m^+, g) \neq \emptyset$, and the existence of $(Y,Z) \in \mathcal{A}(\xi + m, g)$, such that $\mathcal{A}_t(Y_t + m^-, g) \neq \emptyset$, then $\hat{\mathcal{E}}^g_t(\xi + m) = \hat{\mathcal{E}}^g_t(\xi) + m$.*

Proof. 1. Given $(Y^i, Z^i) \in \mathcal{A}(\xi, g)$, for $i = 1, 2$, we have to construct $(\bar{Y}, \bar{Z}) \in \mathcal{A}(\xi, g)$, such that $\bar{Y}_t \leq \min\{Y^1_t, Y^2_t\}$. To this end, we define the stopping time

$$\tau = \inf\{s > t : Y^1_s > Y^2_s\} \wedge T$$

and set $\bar{Y} = Y^1 1_{[0,\tau[} + Y^2 1_{[\tau,T[}$, $\bar{Y}_T = \xi$, and $\bar{Z} = Z^1 1_{[0,\tau]} + Z^2 1_{]\tau,T]}$. Since $Y^1_\tau \geq Y^2_\tau$, Lemma 1.1 yields $(\bar{Y}, \bar{Z}) \in \mathcal{A}(\xi, g)$ and by definition holds $\bar{Y}_t = \min\{Y^1_t, Y^2_t\}$.

2. In view of the first assertion, there exists a sequence $((\tilde{Y}^n, \tilde{Z}^n)) \subset \mathcal{A}(\xi, g)$ such that $(\tilde{Y}_t^n)$ decreases to $\hat{\mathcal{E}}_t^g(\xi)$. Set $Y^n = \tilde{Y}^1 1_{[0,t)} + \tilde{Y}^n 1_{[t,T]}$ and $Z^n = \tilde{Z}^1 1_{[0,t]} + \tilde{Z}^n 1_{(t,T]}$. From Lemma 1.1 follows that $((Y^n, Z^n)) \subset \mathcal{A}(\xi, g)$ and (Y_t^n) decreases to $\hat{\mathcal{E}}_t^g(\xi)$ by construction. Relation (1.2), $Y_0^n = Y_0^1$, and g positive yield

$$\int_0^t Z_s^n dW_s \geq \xi - \int_t^T Z_s^n dW_s + \int_0^T g_s(Y_s^n, Z_s^n)\, ds - Y_0^n \geq -\xi^- - \int_t^T Z_s^n dW_s - Y_0^1.$$

Taking conditional expectation with respect to $\mathcal{F}_t$ and using the supermartingal property of $\int Z^n dW$ yield

$$\int_0^t Z_s^n dW_s \geq -E\left[\xi^- \,\middle|\, \mathcal{F}_t\right] - Y_0^1.$$

Given $\varepsilon > 0$, since $\mathcal{A}(\xi, g) \neq \emptyset$, the sets $B^n = A^n \setminus A^{n-1} \in \mathcal{F}_t$, where $A^n = \{\hat{\mathcal{E}}_t^g(\xi) \geq Y_t^n - \varepsilon\}$ and $A^0 = \emptyset$, form a partition of Ω. Since (Y_t^n) is decreasing, it follows that $Y_t^1 1_{B^n} \geq Y_t^n 1_{B^n}$, for all $n \in \mathbb{N}$. Consequently, by means of Lemma 1.1, $(\bar{Y}, \bar{Z})$, defined as in (1.6), is an element of $\mathcal{A}(\xi, g)$ and $\hat{\mathcal{E}}_t^g(\xi) \geq \bar{Y}_t - \varepsilon$ by construction.

3. Straightforward inspection.

4. The joint convexity of g yields $(\lambda Y + (1-\lambda)Y', \lambda Z + (1-\lambda)Z') \in \mathcal{A}(\lambda\xi + (1-\lambda)\xi', g)$, for all $(Y, Z) \in \mathcal{A}(\xi, g)$, all $(Y', Z') \in \mathcal{A}(\xi', g)$, and all $\lambda \in (0,1)$. Hence, $\lambda\mathcal{A}(\xi, g) + (1-\lambda)\mathcal{A}(\xi', g) \supset \mathcal{A}(\lambda\xi + (1-\lambda)\xi', g)$ and in particular, $\hat{\mathcal{E}}_t^g(\lambda\xi + (1-\lambda)\xi') \leq \lambda\hat{\mathcal{E}}_t^g(\xi) + (1-\lambda)\hat{\mathcal{E}}_t^g(\xi')$.

5. Let us show the cash superadditivity. For $m \in L_t^0$ with $m \geq 0$, given $(Y, Z) \in \mathcal{A}(\xi + m, g)$, and $0 \leq s \leq t' \leq T$, it follows from (1.2) and (Inc) that

$$\begin{aligned} Y_s - m1_{[t,T]}(s) &- \int_s^{t'} g_u(Y_u - m1_{[t,T]}(u), Z_u)du + \int_s^{t'} Z_u dW_u \\ &\geq Y_s - m1_{[t,T]}(s) - \int_s^{t'} g_u(Y_u, Z_u)du + \int_s^{t'} Z_u dW_u \geq Y_{t'} - m1_{[t,T]}(t'). \end{aligned}$$

Hence, $(Y - m1_{[t,T]}, Z) \in \mathcal{A}(\xi, g)$ and thus $\hat{\mathcal{E}}_t^g(\xi + m) - m \geq \hat{\mathcal{E}}_t^g(\xi)$. For the cash subadditivity the same argument yields

$$Y_s + m1_{[t,T]}(s) - \int_s^{t'} g_u(Y_u + m1_{[t,T]}(u), Z_u)du + \int_s^{t'} Z_u dW_u \geq Y_{t'} + m1_{[t,T]}(t'),$$

for all $t \leq s \leq t' \leq T$, and all $(Y, Z) \in \mathcal{A}(\xi, g)$. Given $(\tilde{Y}, \tilde{Z}) \in \mathcal{A}_t(Y_t + m, g)$ our usual pasting argument yields $(\bar{Y}, \bar{Z}) \in \mathcal{A}(\xi + m, g)$, with $Y_t + m = \bar{Y}_t$ and thus $\hat{\mathcal{E}}_t^g(\xi) + m \geq \hat{\mathcal{E}}_t^g(\xi + m)$. The cash additivity in case where g is independent of y follows from $\hat{\mathcal{E}}_t^g(\xi)+m = \hat{\mathcal{E}}_t^g(\xi+m^+)-m^- = \hat{\mathcal{E}}_t^g(\xi+m+m^-)-m^- = \hat{\mathcal{E}}_t^g(\xi+m)$, since (Dec) and (Inc) are simultaneously fulfilled. □

The following Lemma states that under the assumption of a positive generator the value process of a supersolution is a supermartingale. To view supersolutions as supermartingales is one of the key ideas in our approach. We will make ample use of the rich structure supermartingales provide throughout this chapter.

Lemma 1.3. *Let g be a generator fulfilling* (Pos)*, and $\xi \in L^0$ be a terminal condition such that $\xi^- \in L^1$. Let $(Y,Z) \in \mathcal{A}(\xi,g)$. Then $\xi \in L^1$, Y is a supermartingale, Z is unique and Y has the unique decomposition*

$$Y = Y_0 - A + M, \tag{1.7}$$

where M denotes the supermartingale $\int ZdW$ and A is a predictable, increasing, càdlàg process with $A_0 = 0$.

Proof. Relation (1.2), positivity of g, admissibility of Z and $\xi^- \in L^1$ imply $E[|Y_t|] < +\infty$, for all $t \in [0,T]$. Since $-\xi^- \leq \xi \leq Y_T$, we deduce that $\xi \in L^1$. Again, from (1.2), admissibility of Z and positivity of g we derive by taking conditional expectation, that $Y_s \geq E[Y_t \mid \mathcal{F}_s]$. Thus Y is a supermartingale with $Y_t \geq E\left[\xi \mid \mathcal{F}_t\right]$. Relation (1.2) implies further that there exist an increasing and càdlàg process K, with $K_0 = 0$, such that (1.7) holds with $A = \int g(Y,Z)ds + K$. Note that A is optional and therefore predictable due to the Brownian filtration, see [Revuz and Yor, 1999, Corollary V.3.3]. Since Y is a càdlàg supermartingale the Doob-Meyer theorem, see [Protter, 2004, Theorem III.3.13], implies the unique decomposition $Y = Y_0 + \tilde{M} - \tilde{A}$, where $\tilde{M}$ is a local martingale and $\tilde{A}$ is an increasing process which is predictable, and $\tilde{M}_0 = \tilde{A}_0 = 0$. In our filtration every local martingale is continuous, see [Protter, 2004, Corollary IV.3.1, p. 187] and thus $\tilde{A}$ is càdlàg. Hence A and $\tilde{A}$ and M and $\tilde{M}$ are indistinguishable. Moreover, from the predictable representation property of local martingales and from $P(\bigcup_n \{\tau_n = T\}) = 1$, for $\tau^n = \inf\{t \geq 0 | |M_t| \geq n\} \wedge T$, we obtain the $P \otimes dt$-almost sure uniqueness of Z. □

We now prove that for a positive generator $\hat{\mathcal{E}}^g\left(\xi\right)$ is in fact a supermartingale, which, in addition, dominates its right hand limit process. This is crucial for the proof of the existence and uniqueness theorem.

Proposition 1.4. *Let g be a generator fulfilling* (Pos)*, and $\xi \in L^0$ be a terminal condition such that $\xi^- \in L^1$. Suppose that $\mathcal{A}\left(\xi,g\right) \neq \emptyset$, then $\hat{\mathcal{E}}^g\left(\xi\right)$ is a supermartingale,*

$$\mathcal{E}_s^g(\xi) := \lim_{t \downarrow s, t \in \mathbb{Q}} \hat{\mathcal{E}}_t^g(\xi), \qquad \textit{for all } s \in [0,T), \qquad \mathcal{E}_T^g(\xi) := \xi,$$

is a well-defined càdlàg supermartingale, and

$$\hat{\mathcal{E}}_s^g\left(\xi\right) \geq \mathcal{E}_s^g(\xi), \qquad \textit{for all } s \in [0,T]. \tag{1.8}$$

Proof. Note first that $\hat{\mathcal{E}}^g(\xi)$ is adapted by definition. Furthermore, given $(Y,Z) \in \mathcal{A}(\xi,g) \neq \emptyset$, Lemma 1.3 implies $\xi \in L^1$ and $Y_t \geq E\left[\xi \mid \mathcal{F}_t\right]$. Hence $Y_t \geq \hat{\mathcal{E}}_t^g(\xi) \geq E[\xi \mid \mathcal{F}_t]$

and $\hat{\mathcal{E}}^g_t(\xi) \in L^1$. As for the supermartingale property and (1.8), fix $0 \le s \le t \le T$. In view of item 2 of Proposition 1.2, for all $\varepsilon > 0$, there exists $(Y^\varepsilon, Z^\varepsilon) \in \mathcal{A}(\xi, g)$ such that $\hat{\mathcal{E}}^g_s(\xi) \ge Y^\varepsilon_s - \varepsilon$. Due to (1.2) it follows

$$
\begin{aligned}
\hat{\mathcal{E}}^g_t(\xi) \le Y^\varepsilon_t \le Y^\varepsilon_s - \int_s^t g_u(Y^\varepsilon_u, Z^\varepsilon_u)du + \int_s^t Z^\varepsilon_u dW_u \\
\le \hat{\mathcal{E}}^g_s(\xi) - \int_s^t g_u(Y^\varepsilon_u, Z^\varepsilon_u)du + \int_s^t Z^\varepsilon_u dW_u + \varepsilon \le \hat{\mathcal{E}}^g_s(\xi) + \int_s^t Z^\varepsilon_u dW_u + \varepsilon. \quad (1.9)
\end{aligned}
$$

Taking conditional expectation on both sides with respect to $\mathcal{F}_s$ and the supermartingale property of $\int Z^\varepsilon dW$ yields $\hat{\mathcal{E}}^g_s(\xi) \ge E[\hat{\mathcal{E}}^g_t(\xi) \mid \mathcal{F}_s]$, and so $\hat{\mathcal{E}}^g(\xi)$ is a supermartingale. That $\mathcal{E}^g(\xi)$ is well-defined càdlàg supermartingale follows from Karatzas and Shreve [2004, Proposition 1.3.14]. Finally, (1.8) follows directly from (1.9) and the definition of $\mathcal{E}^g(\xi)$. ☐

Remark 1.5. The previous proposition suggests to consider the càdlàg supermartingale $\mathcal{E}^g(\xi)$ as a candidate for the value process of the minimal supersolution. Note further that, if $\mathcal{E}^g(\xi)$ is the value process of the minimal supersolution it is a modification of $\hat{\mathcal{E}}^g(\xi)$ by definition. Hence, in the following, whenever we assume that $\mathcal{E}^g(\xi)$ is minimal, if we make P-almost sure statements, we may write either of them. ♦

The final result of this Section shows that our setup allows to derive various properties that are important in the context of non-linear expectations and dynamic risk measures. In particular, we prove that $\mathcal{E}^g(\xi)$, if it is the value process of the minimal supersolution, fulfills the flow-property and, under the additional assumption $g(y, 0) = 0$, for all $y \in \mathbb{R}$, we show projectivity, with time-consistency as a special case. In the context of BSDE solutions such properties were first established in Peng [1997], for the case of Lipschitz generators. For dynamic risk measures the (strong) time-consistency has been investigated in discrete time in [Cheridito et al., 2006, Föllmer and Penner, 2006] as well as in continuous time in [Bion-Nadal, 2009, Delbaen, 2006], for instance.

Proposition 1.6. *For $t \in [0, T]$, generator g and terminal condition $\xi \in L^0$, it holds*

1. $\hat{\mathcal{E}}^g_{s,T}(\xi) \le \hat{\mathcal{E}}^g_{s,t}(\hat{\mathcal{E}}^g_{t,T}(\xi))$, *for all $0 \le s \le t$. Suppose that $\mathcal{E}^g(\xi)$ is a minimal supersolution, then the* flow-property *holds, that is*
$$\mathcal{E}^g_{s,T}(\xi) = \mathcal{E}^g_{s,t}(\mathcal{E}^g_{t,T}(\xi)), \quad \text{for all } 0 \le s \le t. \tag{1.10}$$

2. *if $g(y, 0) = 0$, for all $y \in \mathbb{R}$, then $\hat{\mathcal{E}}^g_s\left(\hat{\mathcal{E}}^g_t(\xi)\right) \le \hat{\mathcal{E}}^g_s(\xi)$, for all $0 \le s \le t$. Assuming* (Pos), *and supposing that $\mathcal{E}^g(\xi)$ is a minimal supersolution, then $\mathcal{E}^g(\xi)$ is* time-consistent, *that is*
$$\mathcal{E}^g_s(\mathcal{E}^g_t(\xi)) = \mathcal{E}^g_s(\xi), \quad \text{for all } 0 \le s \le t. \tag{1.11}$$

3. *assuming* (Pos), $g(y,0)=0$, *for all* $y\in\mathbb{R}$, *and* $\mathcal{E}^g(\xi)$ *is a minimal supersolution, then the* projectivity *holds, that is*

$$\mathcal{E}_s^g\left(1_A\mathcal{E}_t^g(\xi)\right)=\mathcal{E}_s^g\left(1_A\xi\right),\qquad \text{for all } 0\le s\le t \text{ and } A\in\mathcal{F}_t. \tag{1.12}$$

Proof. 1. Fix $0\le s\le t$. Obviously, $(Y_s,Z_s)_{s\in[0,t]}\in\mathcal{A}_t(\hat{\mathcal{E}}^g_{t,T}(\xi),g)$, for all $(Y,Z)\in\mathcal{A}_T(\xi,g)$. Hence $\hat{\mathcal{E}}^g_{s,t}(\hat{\mathcal{E}}^g_{t,T}(\xi))\le\hat{\mathcal{E}}^g_{s,T}(\xi)$. Suppose now that $\mathcal{E}^g_{\cdot,T}(\xi)$ is a minimal supersolution with corresponding admissible control process $\hat{Z}\in\mathcal{L}$. For all $(Y,Z)\in\mathcal{A}_t(\mathcal{E}^g_{t,T}(\xi),g)$, holds $Y_t\ge\mathcal{E}^g_{t,T}(\xi)$ and, with the same argumentation as in Lemma 1.1, we can paste in a monotone way to show that $(\bar{Y},\bar{Z})\in\mathcal{A}_T(\xi,g)$, where $\bar{Y}=Y1_{[0,t[}+\mathcal{E}^g_{\cdot,T}(\xi)1_{[t,T]}$ and $\bar{Z}=Z1_{[0,t]}+\hat{Z}1_{]t,T]}$. Thus, by definition, $\mathcal{E}^g_{s,t}(\mathcal{E}^g_{t,T}(\xi))\ge\mathcal{E}^g_{s,T}(\xi)$.

2. Given $(Y,Z)\in\mathcal{A}(\xi,g)$, we define $\bar{Y}=Y1_{[0,t[}+\hat{\mathcal{E}}^g_t(\xi)1_{[t,T]}$ and $\bar{Z}=Z1_{[0,t]}$. Since $Y_t\ge\hat{\mathcal{E}}^g_t(\xi)$ and $g(y,0)=0$, it is straightforward to verify that $(\bar{Y},\bar{Z})\in\mathcal{A}(\hat{\mathcal{E}}^g_t(\xi),g)$. From $Y_s\ge\bar{Y}_s$, for all $s\in[0,t]$, follows that $\hat{\mathcal{E}}^g_s(\hat{\mathcal{E}}^g_t(\xi))\le\hat{\mathcal{E}}^g_s(\xi)$, for all $s\in[0,t]$. The case where $\mathcal{E}^g(\xi)$ is a minimal supersolution and Assumption (Pos) holds, follows from (1.12) for $A=\Omega$.

3. Fix $A\in\mathcal{F}_t$. Suppose that $\mathcal{E}^g(\xi)$ is a minimal supersolution with corresponding control process $\hat{Z}$. Given $(Y,Z)\in\mathcal{A}(1_A\mathcal{E}^g_t(\xi),g)$, it follows from Proposition 1.4 below that $Y_t\ge E[1_A\mathcal{E}^g_t(\xi)\mid\mathcal{F}_t]=1_A\mathcal{E}^g_t(\xi)$. Since $g(y,0)=0$, it is straightforward to check that $\tilde{Y}=Y1_{[0,t[}+\mathcal{E}^g_t(\xi)1_A1_{[t,T]}$, and $\tilde{Z}=Z1_{[0,t]}$ is such that $(\tilde{Y},\tilde{Z})\in\mathcal{A}(1_A\mathcal{E}^g_t(\xi),g)$. We can henceforth assume that $Y_s=1_A\mathcal{E}^g_t(\xi)$, for all $s\ge t$. Now, we define $\bar{Y}=Y1_{[0,t[}+\mathcal{E}^g(\xi)1_A1_{[t,T]}$ and $\bar{Z}=Z1_{[0,t]}+\hat{Z}1_A1_{]t,T]}$. For $0\le s<t\le t'\le T$ holds

$$\begin{aligned}\bar{Y}_s-\int_s^{t'}g\left(\bar{Y}_u,\bar{Z}_u\right)du+\int_s^{t'}\bar{Z}_udW_u&\ge Y_t+\left(-\int_t^{t'}g_u\left(\mathcal{E}^g_u(\xi),\hat{Z}_u\right)du+\int_t^{t'}\hat{Z}_udW_u\right)1_A\\&\ge\left(\mathcal{E}^g_t(\xi)-\int_t^{t'}g_u\left(\mathcal{E}^g_u(\xi),\hat{Z}_u\right)du+\int_t^{t'}\hat{Z}_udW_u\right)1_A\ge\mathcal{E}^g_{t'}(\xi)1_A.\end{aligned}$$

Hence, for all $0\le s\le t'\le T$, holds

$$\bar{Y}_s-\int_s^{t'}g\left(\bar{Y}_u,\bar{Z}_u\right)du+\int_s^{t'}\bar{Z}_udW_u\ge Y_{t'}1_{\{t'<t\}}+\mathcal{E}^g_{t'}(\xi)1_A1_{\{t\le t'\}}=\bar{Y}_{t'}$$

and $\bar{Y}_T=1_A\xi$, which implies that $(\bar{Y},\bar{Z})\in\mathcal{A}(1_A\xi,g)$. Since $\bar{Y}_s=Y_s$, for all $s\le t$, we deduce $\mathcal{E}^g_s(1_A\xi)\le\mathcal{E}^g_s(1_A\mathcal{E}_t(\xi))$.

On the other hand, consider $(Y,Z)\in\mathcal{A}(1_A\xi,g)$. From $Y_t\ge E[1_A\xi\mid\mathcal{F}_t]=1_AE[\xi\mid\mathcal{F}_t]$, we obtain $Y_t1_{A^c}\ge0$. Since $\mathcal{E}^g(\xi)$ is a minimal supersolution, it follows that $Y_t\ge\mathcal{E}^g_t(\xi)1_A$. Indeed, let $B=\{Y_t<\mathcal{E}^g_t(\xi)1_A\}$, then $Y_t1_{A^c}\ge0$ implies $B\subset A$. Consequently, by similar arguments as in Lemma 1.1, the processes $\tilde{Y}=\mathcal{E}^g(\xi)(1_{[0,t[}+1_{B^c}1_{[t,T]})+Y1_B1_{[t,T]}$ and $\tilde{Z}=\hat{Z}(1_{[0,t[}+1_{B^c}1_{[t,T]})+Z1_B1_{[t,T]}$ are such that $(\tilde{Y},\tilde{Z})\in$

$\mathcal{A}(\xi, g)$, which implies $P[B] = 0$. It is also straightforward to check that $\tilde{Y} = Y 1_{[0,t[} + \mathcal{E}^g(\xi) 1_A 1_{[t,T]}$ and $\tilde{Z} = Z 1_{[0,t]} + \hat{Z} 1_{(t,T]} 1_A$ are such that $(\tilde{Y}, \tilde{Z}) \in \mathcal{A}(1_A \xi, g)$. Thus we can assume that $Y_t = 1_A \mathcal{E}^g_t(\xi)$. Defining $\bar{Y} = Y 1_{[0,t[} + \mathcal{E}^g_t(\xi) 1_A 1_{[t,T]}$ and $\bar{Z} = Z 1_{[0,t]}$, it holds $(\bar{Y}, \bar{Z}) \in \mathcal{A}(1_A \mathcal{E}^g_t(\xi), g)$. Thus $\mathcal{E}^g_s(1_A \mathcal{E}^g_t(\xi)) \leq \mathcal{E}^g_s(1_A \xi)$, since $\bar{Y}_s = Y_s$, for all $s \leq t$. □

1.3. Existence, Uniqueness and Stability

In this section, we give conditions, which guarantee the existence and uniqueness of a minimal supersolution. We show that the corresponding value process is given by $\mathcal{E}^g(\xi)$. Moreover, we analyze the stability of $\hat{\mathcal{E}}^g(\xi)$ with respect to perturbations of the terminal condition or the generator. In addition to the assumptions (Pos) and (Inc) or (Dec) introduced above, we require convexity of g in the control variable. To that end, we say that a generator g is

(Con) convex, if $g(y, \lambda z + (1-\lambda) z') \leq \lambda g(y, z) + (1-\lambda) g(y, z')$, for all $y \in \mathbb{R}$, all $z, z' \in \mathbb{R}^{1\times d}$ and all $\lambda \in (0,1)$.

1.3.1. Existence and Uniqueness of Minimal Supersolutions

The following theorem on existence and uniqueness of a minimal supersolution is the first main result of this chapter. Note, that we require that $\mathcal{A}(\xi, g) \neq \emptyset$. In the context of finding minimal elements in some set this is quite a standard assumption, see Peng [1999] for an example in the context of minimal supersolutions. However, let us point out that in many applications $\mathcal{A}(\xi, g) \neq \emptyset$ might be guaranteed by specific model assumptions, see for instance the example on utility maximization in Section 1.3.3. It might also be automatically granted under further assumptions, see Cheridito and Stadje [2012], or for instance if the BSDE $Y_t - \int_t^T g_s(Y_s, Z_s)\, ds + \int_t^T Z_s dW_s = \hat{\xi}$ has a solution $(Y, Z) \in \mathcal{S} \times \mathcal{L}$, such that Z is admissible. In the latter case, $\mathcal{A}(\xi, g) \neq \emptyset$, for all $\xi \in L^0$ such that $\xi^- \in L^1$, with $\hat{\xi} \geq \xi$.

Theorem 1.7. *Let g be a generator fulfilling* (Pos)*,* (Con) *and either* (Inc) *or* (Dec) *and $\xi \in L^0$ be a terminal condition, such that $\xi^- \in L^1$. If $\mathcal{A}(\xi, g) \neq \emptyset$, then there exists a unique minimal supersolution $(\hat{Y}, \hat{Z}) \in \mathcal{A}(\xi, g)$. Moreover, $\mathcal{E}^g(\xi)$ is the value process of the minimal supersolution, that is $(\mathcal{E}^g(\xi), \hat{Z}) \in \mathcal{A}(\xi, g)$.*

Note that Remark 1.5 implies that under the assumptions of Theorem 1.7 the process $\mathcal{E}^g(\xi)$ is in fact a modification of $\hat{\mathcal{E}}^g(\xi)$.

Proof. Step 1: Uniqueness. Given $\hat{Z} \in \mathcal{L}$ such that $(\mathcal{E}^g(\xi), Z) \in \mathcal{A}(\xi, g)$, the definition of $\mathcal{E}^g(\xi)$ implies that for any other supersolution $(Y, Z') \in \mathcal{A}(\xi, g)$ holds $\mathcal{E}^g_t(\xi) \leq Y_t$, for all $t \in [0, T]$. The uniqueness of $\hat{Z}$ follows as in Lemma 1.3.

The remainder of the proof provides existence of $\hat{Z} \in \mathcal{L}$ such that $(\mathcal{E}^g(\xi), \hat{Z}) \in \mathcal{A}(\xi, g)$.

Step 2: Construction of an approximating sequence. For any $n, i \in \mathbb{N}$, let $t_i^n = iT/2^n$. There exist $((Y^n, Z^n)) \subset \mathcal{A}(\xi, g)$ such that

$$\hat{\mathcal{E}}^g_{t_i^n}(\xi) \geq Y^n_{t_i^n} - 1/n, \quad \text{for all } n \in \mathbb{N}, \text{ and all } i = 0, \dots, 2^n - 1, \tag{1.13}$$

and

$$Y^n_t \geq Y^{n+1}_t, \quad \text{for all } t \in [0, T], \text{ and all } n \in \mathbb{N}. \tag{1.14}$$

Indeed, by means of Proposition 1.2.2, for each $n \in \mathbb{N}$, we may select a family $((Y^{n,i}, Z^{n,i}))_{i=0,\dots,2^n-1}$ in $\mathcal{A}(\xi, g)$, such that $\hat{\mathcal{E}}^g_{t_i^n}(\xi) \geq Y^{n,i}_{t_i^n} - 1/n$, $i = 0, \dots, 2^n - 1$. We suitably paste this family in order to obtain (1.13). We start with

$$\bar{Y}^{n,0} = Y^{n,0}, \quad \bar{Z}^{n,0} = Z^{n,0},$$

and continue by recursively setting, for $i = 1, \dots, 2^n - 1$,

$$\begin{aligned}
\bar{Y}^{n,i} &= \bar{Y}^{n,i-1} 1_{[0,\tau_i^n[} + Y^{n,i} 1_{[\tau_i^n, T[}, \quad \bar{Y}^{n,i}_T = \xi, \\
\bar{Z}^{n,i} &= \bar{Z}^{n,i-1} 1_{[0,\tau_i^n]} + Z^{n,i} 1_{]\tau_i^n, T]},
\end{aligned}$$

where τ_i^n are stopping times given by $\tau_i^n = \inf\{t > t_i^n : \bar{Y}^{n,i-1}_t > Y^{n,i}_t\} \wedge T$. From the definition of these stopping times and Lemma 1.1 follows that the pairs $(\bar{Y}^{n,i}, \bar{Z}^{n,i})$, $i = 0, \dots, 2^n - 1$, are elements of $\mathcal{A}(\xi, g)$. Hence, the sequence

$$((Y^n, Z^n) := (\bar{Y}^{n,2^n-1}, \bar{Z}^{n,2^n-1}))$$

fulfills (1.13) by construction. Note that $((Y^n, Z^n))$ is not necessarily monotone in the sense of (1.14). However, this can be achieved by pasting similarly. More precisely, we choose

$$\bar{Y}^1 = Y^1, \quad \bar{Z}^1 = Z^1,$$

and continue by recursively setting, for $n \in \mathbb{N}$,

$$\begin{aligned}
\bar{Y}^n &= \sum_{i=0}^{2^n-1} \left(Y^n 1_{[t_i^n, \tau_i^n[} + \bar{Y}^{n-1} 1_{[\tau_i^n, t_{i+1}^n[} \right), \quad \bar{Y}^n_T = \xi, \\
\bar{Z}^n &= \sum_{i=0}^{2^n-1} \left(Z^n 1_{]t_i^n, \tau_i^n]} + \bar{Z}^{n-1} 1_{]\tau_i^n, t_{i+1}^n]} \right),
\end{aligned}$$

where τ_i^n are stopping times given by $\tau_i^n = \inf\{t > t_i^n : Y^n_t > \bar{Y}^{n-1}_t\} \wedge t_{i+1}^n$, for $i = 0, \dots, 2^n - 1$. By construction $((\bar{Y}^n, \bar{Z}^n))$ fulfills both (1.13) and (1.14), and $((\bar{Y}^n, \bar{Z}^n)) \subset \mathcal{A}(\xi, g)$ with Lemma 1.1.

Step 3: Bound on $\int Z^n dW$. We now take the sequence $((Y^n, Z^n))$ fulfilling (1.13) and (1.14) and provide an inequality which will enable us to use compactness arguments for

(Z^n) later in the proof. More precisely, we argue that, for all $n \in \mathbb{N}$, holds

$$\left| \int_0^t Z_s^n dW_s \right| \leq B_t^n := \left|Y_t^1\right| + E\left[\xi^- \;\middle|\; \mathcal{F}_t\right] + E\left[\xi^-\right] + \left|Y_0^1\right| + A_t^n, \tag{1.15}$$

for all $t \in [0,T]$, where A_t^n is the positive increasing process defined in Lemma 1.3. Moreover, it holds

$$E[A_T^n] \leq Y_0^1 - E[\xi].$$

Indeed, by the same arguments as in the proof of Lemma 1.2.2, recall $Y_0^n \leq Y_0^1$, follows

$$\int_0^t Z_s^n dW_s \geq -E\left[\xi^- \;\middle|\; \mathcal{F}_t\right] - Y_0^1. \tag{1.16}$$

On the other hand, from $Y_t^n \leq Y_t^1$ and $-Y_0^n \leq E[\xi^-]$, recall Lemma 1.3, it follows

$$\int_0^t Z_s^n dW_s \leq Y_t^1 + A_t^n - Y_0^n \leq Y_t^1 + A_t^n + E[\xi^-]. \tag{1.17}$$

Combining (1.17) and (1.16) yields (1.15). The L^1 bound on A^n follows from $Y_0^n - A_T^n + \int_0^T Z_s^n dW_s = \xi$, $Y_0^1 \geq Y_0^n$, and the supermartingal property of $\int Z^n dW$.

Note that if $(B_T^{n,*})$ in (1.15) were bounded in L^1, then, by means of the Burkholder-Davis-Gundy inequality, (Z^n) would be a bounded sequence in $\mathcal{L}^1$ and we could apply [Delbaen and Schachermayer, 1996, Theorem A] to find a sequence in the asymptotic convex hull of (Z^n), which converges in $\mathcal{L}^1$ and $P \otimes dt$-almost surely along some localizing sequence of stopping times to some limit $Z \in \mathcal{L}^1$. Here, even if $(A_T^{n,*}) = (A_T^n)$ is uniformly bounded, this is however not necessarily the case for $Y_T^{1,*}$ and $(E[\xi^- \mid \mathcal{F}_\cdot])_T^*$, and this is the reason why we introduce the following localization.

Step 4: First localization. Due to our Brownian setting and since $\xi^- \in L^1$, we know that the martingale $E[\xi^- \mid \mathcal{F}_\cdot]$, has a continuous version, see [Revuz and Yor, 1999, Theorem V.3.5]. Moreover, Y^1 is a càdlàg supermartingale and thus we may take the localising sequence

$$\sigma_k = \inf\left\{t > 0 : \left|Y_t^1\right| + E\left[\xi^- \;\middle|\; \mathcal{F}_t\right] > k\right\} \wedge T, \quad k \in \mathbb{N}, \tag{1.18}$$

which is independent of $n \in \mathbb{N}$. For a fixed $k \in \mathbb{N}$, Inequality (1.15) yields

$$\left(\int Z^n 1_{[0,\sigma_k]} dW\right)_T^* \leq B^{k,n}, \qquad \text{for all } n \in \mathbb{N}, \tag{1.19}$$

where $B^{k,n} = \left|Y_0^1\right| + E[\xi^-] + k + A_T^n$. Due to $E[A_T^n] \leq Y_0^1 - E[\xi]$ we have

$$\sup_{n\in\mathbb{N}} E\left[B^{k,n}\right] < \infty. \tag{1.20}$$

Since $(B^{k,n})_{n\in\mathbb{N}}$ is a sequence of positive random variables we may apply [Delbaen and Schachermayer, 1994, Lemma A1.1]. It provides a sequence $(\tilde{B}^{k,n})_{n\in\mathbb{N}}$ in the asymptotic convex hull of $(B^{k,n})_{n\in\mathbb{N}}$, which converges almost surely to a random variable $\tilde{B}^k \geq 0$. The $\tilde{B}^{k,n}$ inherit the integrability of the $B^{k,n}$ and we can conclude with Fatou's lemma that

$$E[\tilde{B}^k] < \infty. \tag{1.21}$$

Let $\tilde{Z}^{k,n}$ be the convex combination of (Z^n) corresponding to $\tilde{B}^{k,n}$ so that

$$\left(\int \tilde{Z}^{k,n} 1_{[0,\sigma_k]} dW\right)_T^* \leq \tilde{B}^{k,n}, \qquad \text{for all } n \in \mathbb{N}. \tag{1.22}$$

Step 5: Second localization. The next two steps follow some known compactness arguments, which, in the case of $\mathcal{L}^1$, can be found in [Delbaen and Schachermayer, 1996]. For the sake of completeness we develop the argumentation. Given an $m \in \mathbb{N}$, we start by taking a fast subsequence $(\tilde{B}^{k,m,n})_{n\in\mathbb{N}}$ of $(\tilde{B}^{k,n})_{n\in\mathbb{N}}$ converging in probability to $\tilde{B}^k$. More precisely, we choose $(\tilde{B}^{k,m,n})_{n\in\mathbb{N}}$ such that

$$P\left[\left|\tilde{B}^{k,m,n} - \tilde{B}^k\right| \geq 1\right] \leq \frac{2^{-n}}{m}. \tag{1.23}$$

Consider now the stopping time $\tau^{k,m}$ given by

$$\tau^{k,m} = \inf\left\{t \geq 0 : \left(\int \tilde{Z}^{k,m,n} 1_{[0,\sigma_k]} dW\right)_t^* \geq m, \text{ for some } n \in \mathbb{N}\right\} \wedge T, \tag{1.24}$$

where the sequence $(\tilde{Z}^{k,m,n}1_{[0,\sigma_k]})_{n\in\mathbb{N}}$ is the subsequence of $(\tilde{Z}^{k,n}1_{[0,\sigma_k]})_{n\in\mathbb{N}}$ corresponding to the fast subsequence $(\tilde{B}^{k,m,n})_{n\in\mathbb{N}}$. The definition of $\tau^{k,m}$ as well as the Burkholder-Davis-Gundy inequality imply that the sequence of processes $(\tilde{Z}^{k,m,n}1_{[0,\sigma_k]}1_{[0,\tau^{k,m}]})_{n\in\mathbb{N}}$ is bounded in $\mathcal{L}^2$. The Alaoglu-Bourbaki theorem and the Eberlein-Šmulian theorem in the Banach space $\mathcal{L}^2$ imply the existence of $\hat{Z}^{k,m} \in \mathcal{L}^2$, such that, up to a subsequence, $(\tilde{Z}^{k,m,n}1_{[0,\sigma_k]}1_{[0,\tau^{k,m}]})_{n\in\mathbb{N}}$ converges weakly to $\hat{Z}^{k,m}$. As a consequence of the Hahn-Banach theorem, there exists a sequence in the asymptotic convex hull of $(\tilde{Z}^{k,m,n}1_{[0,\sigma_k]}1_{[0,\tau^{k,m}]})_{n\in\mathbb{N}}$, again denoted with $(\tilde{Z}^{k,m,n}1_{[0,\sigma_k]}1_{[0,\tau^{k,m}]})_{n\in\mathbb{N}}$, which converges in $\mathcal{L}^2$ to $\hat{Z}^{k,m}$. By taking another subsequence we also have the $P\otimes dt$-almost sure convergence.

Step 6: $(\tau^{k,m})_{m\in\mathbb{N}}$ is a localizing sequence of stopping times. We estimate as follows

$$\begin{aligned}
P\left[\tau^{k,m}=T\right] &= P\left[\left(\int \tilde{Z}^{k,m,n}1_{[0,\sigma_k]}dW\right)^*_T < m \text{ , for all } n\in\mathbb{N}\right] \\
&\geq 1-P\left[\tilde{B}^{k,m,n}\geq m \text{ , for some } n\in\mathbb{N}\right] \\
&\geq 1-P\left[\left\{\left|\tilde{B}^{k,m.n}-\tilde{B}^k\right|\geq 1 \text{ , for some } n\in\mathbb{N}\right\}\cup\left\{\tilde{B}^k>m-1\right\}\right] \\
&\geq 1-\sum_n P\left[\left|\tilde{B}^{k,m,n}-\tilde{B}^k\right|\geq 1\right]-P\left[\tilde{B}^k>m-1\right] \\
&\geq 1-\frac{1}{m}-\frac{E\left[\tilde{B}^k+1\right]}{m}\xrightarrow[m\to\infty]{}1,
\end{aligned}$$

where we used (1.22) in the second line and (1.23), the Markov inequality and the fact that $E[\tilde{B}^k]<\infty$ in the last one.

Step 7: Construction of the candidate $\hat{Z}$. For given $k,m>0$, we constructed in Step 5 the process $\hat{Z}^{k,m}$ as the $\mathcal{L}^2$ and $P\otimes dt$-almost sure limit of a sequence in the asymptotic convex hull of $(\tilde{Z}^{k,m,n}1_{[0,\sigma_k]}1_{[0,\tau^{k,m}]})_{n\in\mathbb{N}}$. With $(\tilde{B}^{k,m,n})_{n\in\mathbb{N}}$ we denote the corresponding subsequence of convex combinations of $(\tilde{B}^{k,m,n})_{n\in\mathbb{N}}$ and note that $(\int\tilde{Z}^{k,m,n}1_{[0,\sigma_k]}dW)^*_T\leq\tilde{B}^{k,m,n}$, for all $n\in\mathbb{N}$, as in (1.22). Hence, by the same procedure as in Step 5, we can find, for $m'>m$, a fast subsequence $(\tilde{Z}^{k,m',n}1_{[0,\sigma_k]})_{n\in\mathbb{N}}$ in the asymptotic convex hull of $(\tilde{Z}^{k,m,n}1_{[0,\sigma_k]})_{n\in\mathbb{N}}$ and a $\hat{Z}^{k,m'}\in\mathcal{L}^2$ such that $(\tilde{Z}^{k,m',n}1_{[0,\sigma_k]}1_{[0,\tau^{k,m'}]})_{n\in\mathbb{N}}$ converges in $\mathcal{L}^2$ and $P\otimes dt$-almost surely to $\hat{Z}^{k,m'}$. We iterate this procedure and define $(\tilde{Z}^{k,n})_{n\in\mathbb{N}}$ as the diagonal sequence $\tilde{Z}^{k,n}=\tilde{Z}^{k,n,n}$ and $\hat{Z}^k$ as

$$Z^k_0=0,\quad \hat{Z}^k=\sum_{m=1}^{\infty}\hat{Z}^{k,m}1_{]\tau^{k,m-1},\tau^{k,m}]}. \tag{1.25}$$

From $\hat{Z}^{k,m'}1_{[0,\tau^{k,m}]}=\hat{Z}^{k,m}$, for $m'>m$, follows that $(\tilde{Z}^{k,n}1_{[0,\sigma_k]}1_{[0,\tau^{k,n}]})_{n\in\mathbb{N}}$ converges in $\mathcal{L}^2$ and $P\otimes dt$-almost surely to $\hat{Z}^k$. With the sequence $(\tilde{Z}^{k,n})_{n\in\mathbb{N}}$ and the process $\tilde{Z}^k$ at hand, we now diagonalize our program above with respect to k and n. As before, we get a diagonal sequence $\tilde{Z}^n=\tilde{Z}^{n,n}$, and a process $\hat{Z}$ given by

$$\hat{Z}_0=0,\quad \hat{Z}=\sum_{k=1}^{\infty}1_{]\sigma_{k-1},\sigma_k]}\hat{Z}^k, \tag{1.26}$$

such that

$$\tilde{Z}^n1_{[0,\tau_n]}\xrightarrow[n\to\infty]{P\otimes dt\text{-almost surely}}\hat{Z}, \tag{1.27}$$

for $\tau_n=\sigma_n\wedge\tau^{n,n}$, where σ_n and $\tau^{n,n}$ are as in (1.18) and (1.24), respectively. For later reference, note that by construction holds $\tilde{Z}^{k',m}1_{[0,\sigma_k]}1_{[0,\tau^{k,m}]}=\tilde{Z}^{k,m}$, as soon as $k'\geq k$ and also $\tilde{Z}1_{[0,\sigma_k]}1_{[0,\tau^{k,m}]}=\tilde{Z}^{k,m}$. Likewise $(\tilde{Z}^n1_{[0,\tau_l]})_{n\in\mathbb{N}}$ converges in $\mathcal{L}^2$ and $P\otimes dt$-almost surely to $\hat{Z}^{l,l}$. This yields, via the Burkholder-Davis-Gundy inequality, up to a subsequence,

$$\int_0^{t\wedge\tau_l} \tilde{Z}^n_s dW_s \xrightarrow[n\to+\infty]{} \int_0^{t\wedge\tau_l} \hat{Z}_s dW_s, \quad \text{for all } t\in[0,T],\ P\text{-almost surely.} \tag{1.28}$$

Hence, diagonalizing yields

$$\int_0^{t} \tilde{Z}^n_s dW_s \xrightarrow[n\to+\infty]{} \int_0^{t} \hat{Z}_s dW_s, \quad \text{for all } t\in[0,T],\ P\text{-almost surely.} \tag{1.29}$$

Step 8: Monotone convergence to $\mathcal{E}^g(\xi)$. Let $\tilde{Y}_t = \lim_n Y^n_t$, for $t\in[0,T]$, be the pointwise monotone limit of the sequence (Y^n). By monotone convergence $\tilde{Y}$ is a supermartingale and, since our filtration is right-continuous, by standard arguments we may define the càdlàg supermartingale $\hat{Y}$ by setting $\hat{Y}_t = \lim_{s\downarrow t, s\in\mathbb{Q}} \tilde{Y}_s$, for all $t\in[0,T)$, and $\hat{Y}_T=\xi$. By construction $\tilde{Y}_{t^i_n} = \hat{\mathcal{E}}^g_{t^i_n}(\xi)$. Hence, $\hat{Y}_t = \mathcal{E}^g_t(\xi)$, for all $t\in[0,T]$, and

$$Y^n_t \geq \tilde{Y}_t \geq \hat{\mathcal{E}}^g_t(\xi) \geq \mathcal{E}^g_t(\xi) \geq E\left[\xi \,\middle|\, \mathcal{F}_t\right], \tag{1.30}$$

where the third inequality follows from Proposition 1.4. Now, the process $\mathcal{E}^g(\xi)$ is the natural candidate for the value process of the minimal supersolution for two reasons. It is càdlàg and it is dominated by $\hat{\mathcal{E}}^g(\xi)$ as (1.30) shows. However, it is not clear a priori that the sequence (Y^n) converges to $\mathcal{E}^g(\xi)$ in some suitable sense. Taking into account the additional structure provided by the supermartingale property of the Y^n we can prove nonetheless

$$\mathcal{E}^g(\xi) = \hat{Y} = \lim_{n\to\infty} Y^n, \quad P\otimes dt\text{-almost surely.} \tag{1.31}$$

Indeed, recall the decomposition $Y^n_t = Y^n_0 - A^n_t + M^n_t$, for all $t\in[0,T]$, and the L^1-bound on (A^n_T) given in Step 3. We now consider the sequence $((\tilde{Y}^n,\tilde{Z}^n))$ in the asymptotic convex hull of (Y^n,Z^n), which corresponds to the sequence $(\tilde{Z}^n)$ constructed in Step 7. From the decomposition of the Y^n, see Lemma 1.3, we obtain that $\tilde{Y}^n_t = \tilde{Y}^n_0 - \tilde{A}^n_t + \tilde{M}^n_t$, for all $t\in[0,T]$. Moreover, the $(\tilde{A}^n_T)$ inherit the L^1-bound given in Step 3. By means of Helly's theorem, see Lemma 1.25, we obtain the existence of a subsequence $(\tilde{A}^n)$ in the asymptotic convex hull of $(\tilde{A}^n)$ and the existence of an increasing positive integrable process $\tilde{A}$, such that $\lim_{k\to\infty}\tilde{A}^{n_k}_t = \tilde{A}_t$, for all $t\in[0,T]$, P-almost surely. Consequently, by monotonicity of (Y^n) and (1.29), $\tilde{Y}_t = \tilde{Y}_0 - \tilde{A}_t + \tilde{M}_t$, for all $t\in[0,T]$. Hence, the jumps of $\tilde{Y}$ are given by the countably many jumps of the increasing process $\tilde{A}$, which implies

$$\hat{Y}_t = \tilde{Y}_0 - \lim_{s\downarrow t, s\in\mathbb{Q}} \tilde{A}_s + \tilde{M}_t, \qquad \text{for all } t\in[0,T), \qquad \hat{Y}_T = \xi.$$

Moreover, the jumptimes of the càdlàg process $\hat{Y}$ are exhausted by a sequence of stopping times $(\rho^j)\subset\mathcal{T}$, which coincide with the jumptimes of $\tilde{A}$. Therefore, $\hat{Y}=\tilde{Y}$,

$P \otimes dt$-almost surely, which implies (1.31).

Step 9: Verification. Let us now show that $(\mathcal{E}^g(\xi), \hat{Z}) \in \mathcal{A}(\xi, g)$, which, by means of (1.30), would end the proof. We start with the verification of (1.2) under the Assumption (Inc). Due to (1.31) there exists a Lebesgue nullset $\mathcal{I} \subset [0,T]$, such that $\mathcal{E}^g_t(\xi) = \lim_{n\to\infty} Y^n_t$, P-almost surely, for all $t \in \mathcal{I}^c$. Let $s, t \in \mathcal{I}^c$ with $s \leq t$. By using (1.29), the $P \otimes dt$-almost sure convergence of $\tilde{Z}^n 1_{[0,\tau^n]}$ to $\hat{Z}$, and Fatou's lemma we obtain

$$\mathcal{E}^g_s - \int_s^t g_u(\mathcal{E}^g_u, \hat{Z}_u)du + \int_s^t \hat{Z}_u dW_u \geq \limsup_n \left(\tilde{Y}^n_s - \int_s^t g_u(\mathcal{E}^g_u, \tilde{Z}^n_u 1_{[0,\tau^n]}(u))du + \int_s^t \tilde{Z}^n_u dW_u \right), \quad (1.32)$$

where $\tilde{Y}^n$ denotes the convex combination of (Y^n) corresponding to $\tilde{Z}^n$. We denote by $\lambda^{(n)}_i$, $n \leq i \leq M^{(n)}$, $\lambda^{(n)}_i \geq 0, \sum_i \lambda^{(n)}_i = 1$ the convex weights of $\tilde{Z}^n$. Since our generator fullills (Con), and since, for n large enough, we have $\tilde{Z}^n_u 1_{[0,\tau^n]}(u) = \tilde{Z}^n_u$, for all $s \leq u \leq t$, we may further estimate the above by

$$\mathcal{E}^g_s - \int_s^t g_u(\mathcal{E}^g_u, \hat{Z}_u)du + \int_s^t \hat{Z}_u dW_u \geq \limsup_n \sum_{i=n}^{M^{(n)}} \lambda^{(n)}_i \left(Y^i_s - \int_s^t g_u(\mathcal{E}^g_u, Z^i_u)du + \int_s^t Z^i_u dW_u \right).$$

Since $Y^i_t \geq \hat{\mathcal{E}}^g_t(\xi) \geq \mathcal{E}^g_t(\xi)$, for all $t \in [0,T]$, and $i \in \mathbb{N}$, we use (Inc) and the fact that the (Y^n, Z^n) are supersolutions to conclude

$$\begin{aligned}\mathcal{E}^g_s - \int_s^t g_u(\mathcal{E}^g_u, \hat{Z}_u)du + \int_s^t \hat{Z}_u dW_u &\geq \limsup_n \sum_{i=n}^{M^{(n)}} \lambda^{(n)}_i \left(Y^i_s - \int_s^t g_u(Y^i_u, Z^i_u)du + \int_s^t Z^i_u dW_u \right) \\ &\geq \limsup_n \sum_{i=n}^{M^{(n)}} \lambda^{(n)}_i Y^i_t = \limsup_n \tilde{Y}^n_t = \limsup_n Y^n_t = \mathcal{E}^g_t. \quad (1.33)\end{aligned}$$

As for the case of $s, t \in \mathcal{I}$, with $s \leq t$, we approximate them both from the right by some sequences $(s^n) \subset \mathcal{I}^c$ and $(t^n) \subset \mathcal{I}^c$, such that $s^n \downarrow s$, $t^n \downarrow t$, $s^n \leq t^n$. For each s^n and t^n holds (1.33). Passing to the limit by using the right-continuity of $\mathcal{E}^g$ and the continuity of $-\int g(\mathcal{E}^g, \hat{Z})du + \int \hat{Z}dW$ we deduce that (1.33), holds for all $s, t \in [0,T]$ with $s \leq t$.

It remains to show admissibility of $\hat{Z}$. By means of (1.33), (1.30), and positivity of g it holds

$$\int_0^t \hat{Z}_s dW_s \geq E\left[\xi \,\middle|\, \mathcal{F}_t\right] - \mathcal{E}_0. \tag{1.34}$$

Being bounded from below by a martingale, the continuous local martingale $\int \hat{Z} dW$ is by Fatou's lemma a supermartingale and thus $\hat{Z}$ is admissible. Hence, the proof under Assumptions (Pos), (Con) and (Inc) is completed.

The proof under (Dec) replacing (Inc) only differs in the verification of (1.2). Indeed, instead of only approximating $\hat{Z}$ in the Lebesgue integral we approximate $\mathcal{E}^g(\xi)$ $P \otimes dt$-almost surely with the sequence (Y^n) as well, that is (1.32) becomes, by means of (1.31) and Fatou's lemma,

$$\begin{aligned}\mathcal{E}^g_s - \int_s^t g_u(\mathcal{E}^g_u, \hat{Z}_u) du + \int_s^t \hat{Z}_u dW_u & \\ \geq \limsup_n & \left(\tilde{Y}^n_s - \int_s^t g_u(Y^n_u, \tilde{Z}^n_u 1_{[0,\tau^n]}(u)) du + \int_s^t \tilde{Z}^n_u dW_u \right).\end{aligned}$$

This entails, by monotonicity of the sequence (Y^n) and the fact that the convex combinations in $\tilde{Z}^n$ consist of elements of (Z^i) with index greater or equal than n, that we may write $-\int_s^t g_u(Y^n_u, Z^i_u) du \geq -\int_s^t g_u(Y^i_u, Z^i_u) du$ in (1.33) and this ends the proof. □

Theorem 1.7 ensures the existence and uniqueness of the minimal supersolution which is càdlàg. The following proposition provides a condition under which $\mathcal{E}^g(\xi)$ is in fact continuous.

Proposition 1.8. *Let g be a generator fulfilling* (Pos), (Con) *and either* (Inc) *or* (Dec) *and $\xi \in L^0$ be a terminal condition, such that $\xi^- \in L^1$. Suppose that $\mathcal{A}(\xi, g) \neq \emptyset$. Assume that for any $\zeta \in L^\infty(\mathcal{F}_\tau)$, $\tau \in \mathcal{T}$, there exist $Y \in \mathcal{S}$ and an admissible $Z \in \mathcal{L}$, which solve the backward stochastic differential equation*

$$Y_t - \int_t^\tau g_s(Y_s, Z_s) ds + \int_t^\tau Z_s dW_s = \zeta, \qquad \textit{for all } t \in [0, \tau].$$

Then $\mathcal{E}^g(\xi)$ is continuous.

Proof. In view of Theorem 1.7, there exists $\hat{Z} \in \mathcal{L}$ such that $(\mathcal{E}^g, \hat{Z}) \in \mathcal{A}(\xi, g)$. Hence, $\mathcal{E}^g$ can only have negative jumps. Assume that $\mathcal{E}^g$ has a negative jump, that is $P[\tau \leq T] > 0$, for the stopping time $\tau = \inf\{t > 0 : \Delta\mathcal{E}^g_t < 0\}$. We then fix m big enough such that the stopping time $\tau^m = \inf\{t > 0 : |\mathcal{E}^g_t| > m\} \wedge \tau$ satisfies $P[\{-m < \Delta\mathcal{E}^g_{\tau^m} < 0\} \cap \{\tau^m = \tau\}] > 0$. Since $\mathcal{E}^g$ is continuous on $[0, \tau[$ and $\mathcal{E}^g$ has only negative jumps, $\mathcal{E}^g_{\tau^m} \vee -m \in L^\infty(\mathcal{F}_{\tau^m})$. By assumption there exist $\bar{Y} \in \mathcal{S}$ and an admissible $\bar{Z} \in \mathcal{L}$

such that

$$\bar{Y}_s + \int_s^{\tau^m} g_u\left(\bar{Y}_u, \bar{Z}_u\right) - \int_s^{\tau^m} \bar{Z}_u dW_u = \mathcal{E}^g_{\tau^m} \vee -m, \qquad \text{for all } s \in [0, \tau^m].$$

Similar to Lemma 1.1, we derive $(\bar{Y} 1_{[0,\tau^m[} + \mathcal{E}^g 1_{[\tau^m,T]}, \bar{Z} 1_{[0,\tau^m]} + \hat{Z} 1_{]\tau^m,T]}) \in \mathcal{A}(\xi, g)$. Hence, by optimality of $\mathcal{E}^g$ in $\mathcal{A}(\xi, g)$ holds $\mathcal{E}^g \leq \bar{Y} 1_{[0,\tau^m[} + \mathcal{E}^g 1_{[\tau^m,T]}$. Moreover, we have

$$\mathcal{E}^g_{\tau^m -} > \mathcal{E}^g_{\tau^m} \vee -m = \bar{Y}_{\tau^m} = \bar{Y}_{\tau^m -} \quad \text{on the set } \{-m < \Delta \mathcal{E}^g_{\tau^m} < 0\} \cap \{\tau^m = \tau\}.$$

Hence, for the stopping time $\hat{\tau} = \inf\{t > 0 : \mathcal{E}^g_t > \bar{Y}_t\} \wedge \tau^m$ we deduce $P[\hat{\tau} < \tau^m] > 0$, since the processes $\mathcal{E}^g$ and $\bar{Y}$ are continuous on $[0, \tau^m[$. But then $\mathcal{E}^g \not\leq \bar{Y}$ on $[0, \tau^m[$, which is a contradiction. □

Under the assumptions of Theorem 1.7, $\mathcal{E}^g$ is the value process of the minimal supersolution with a control process $\hat{Z}$ in $\mathcal{L}$ which defines a supermartingale. Next we address the question whether and under which conditions some stronger assumptions on the control process can be obtained. More precisely, that the corresponding control process $\hat{Z}$ belongs to some $\mathcal{L}^p$, for $p \geq 1$, and therefore defines a true martingale instead of a supermartingale. Defining

$$\mathcal{A}^p(\xi, g) := \{(Y, Z) \in \mathcal{A}(\xi, g) : Z \in \mathcal{L}^p\}, \tag{1.35}$$

this means that $(\mathcal{E}^g(\xi), \hat{Z}) \in \mathcal{A}^p(\xi, g)$. Peng [1999] provides a positive answer to that question in the case where $p = 2$, the terminal condition $\xi \in L^2$ and the generator is not necessarily positive but Lipschitz. Compare also with Cheridito and Stadje [2012] for supersolutions of BSDEs where the control process is in BMO, if the terminal condition is a bounded lower semicontinuous function of the Brownian motion and the generator is convex in z and Lipschitz and increasing in y. Here, we provide an answer to the case where $p = 1$ in the context of Section 1.2. Given a terminal condition ξ, obtaining $\mathcal{E}^g(\xi)$ as a minimal solution with a control process within $\mathcal{L}^1$ comes at two costs. Indeed, a stronger integrability condition on the terminal value is required, that is, we impose that $(E[\xi^- \mid \mathcal{F}_\cdot])^*_T \in L^1$. As for the second cost, $\mathcal{A}^1(\xi, g) \neq \emptyset$ is also required, which, in view of $\mathcal{A}^1(\xi, g) \subset \mathcal{A}(\xi, g)$, is also a stronger assumption.

Theorem 1.9. *Suppose that the generator g fulfills* (Pos)*,* (Con) *and either* (Inc) *or* (Dec)*. Let $\xi \in L^0$ be a terminal condition, such that $(E[\xi^- \mid \mathcal{F}_\cdot])^*_T \in L^1$. If $\mathcal{A}^1(\xi, g) \neq \emptyset$, then there exists a unique minimal supersolution $(\hat{Y}, \hat{Z}) \in \mathcal{A}^1(\xi, g)$. Moreover, $\mathcal{E}^g(\xi)$ is the value process of the minimal supersolution, that is $(\mathcal{E}^g(\xi), \hat{Z}) \in \mathcal{A}^1(\xi, g)$.*

Remark 1.10. As in Section 1.2, note that for $(Y, Z) \in \mathcal{A}^1(\xi, g)$, the value process Y is a supermartingale with terminal value greater or equal than ξ. Moreover, we have $Y^*_T \in L^1$. Indeed, by using the decomposition (1.7), we derive $Y^*_t \leq |Y_0| + A_T + \left(\int Z dW\right)^*_T$.

We further have $A_T \leq Y_0+\int_0^T Z_s dW_s - \xi$ and thus $E\left[|A_T|\right] \leq Y_0+E\left[\xi^-\right]$. Consequently

$$E\left[Y_T^*\right] \leq |Y_0| + E\left[\xi^-\right] + Y_0 + E\left[\left(\left(\int ZdW\right)_T^*\right)\right] < \infty.$$

♦

Proof (of Theorem 1.9). Since $\mathcal{A}^1(\xi,g) \subset \mathcal{A}(\xi,g)$, the assumption $\mathcal{A}^1(\xi,g) \neq \emptyset$ implies the existence of $\hat{Z} \in \mathcal{L}$ such that $(\mathcal{E}^g(\xi), \hat{Z}) \in \mathcal{A}(\xi,g)$. We are left to show that $\hat{Z} \in \mathcal{L}^1$. Since $\mathcal{A}^1(\xi,g) \neq \emptyset$, we can suppose in the proof of Theorem 1.7 that $(Y^1, Z^1) \in \mathcal{A}^1(\xi,g)$. Since (1.15) holds for $(\mathcal{E}^g(\xi), \hat{Z})$, instead of (Y^n, Z^n), we have

$$\left(\int \hat{Z}dW\right)_T^* \leq \left|Y_0^1\right| + E\left[\xi^-\right] + \hat{A}_T + (Y^1)_T^* + \left(E\left[\xi^- \;\middle|\; \mathcal{F}_\cdot\right]\right)_T^*, \quad \text{for all } n \in \mathbb{N}, \tag{1.36}$$

where $0 \leq E[\hat{A}_T] \leq E[\xi] - Y_0^1$. Since $(E[\xi^- \mid \mathcal{F}_\cdot])_T^* \in L^1$, by means of Remark 1.10, the right hand side of (1.36), is in L^1. Thus, by means of the Burkholder-Davis-Gundy inequality, $\hat{Z}$ belongs to $\mathcal{L}^1$. □

1.3.2. Stability Results

In this section we address the stability of $\hat{\mathcal{E}}^g(\cdot)$ with respect to perturbations of the terminal condition or the generator. First we show that the functional $\hat{\mathcal{E}}_0^g$ is not only defined on the same domain as the usual expectation, but also shares some of its main properties, such as Fatou's lemma as well as a monotone convergence theorem.

Theorem 1.11. *Suppose that the generator g fulfills* (Pos), (Con) *and either* (Inc) *or* (Dec). *Let (ξ^n) be a sequence in L^0, such that $\xi^n \geq \eta$, for all $n \in \mathbb{N}$, where $\eta \in L^1$.*

- *Monotone convergence: If (ξ^n) is increasing P-almost surely to $\xi \in L^0$, then $\hat{\mathcal{E}}_0^g(\xi) = \lim_n \hat{\mathcal{E}}_0^g(\xi^n)$.*
- *Fatou's lemma: $\hat{\mathcal{E}}_0^g(\liminf_n \xi^n) \leq \liminf_n \hat{\mathcal{E}}_0^g(\xi^n)$.*

Proof. Monotone convergence: From Proposition 1.2 and by monotonicity, it follows that $\hat{\mathcal{E}}^g(\xi^n) \leq \hat{\mathcal{E}}^g(\xi^{n+1}) \leq \cdots \leq \hat{\mathcal{E}}^g(\xi)$. Hence, we may define $\hat{Y}_0 = \lim_n \hat{\mathcal{E}}_0^g(\xi^n)$. Note that $\hat{Y}_0 \leq \hat{\mathcal{E}}_0^g(\xi)$. If $\hat{Y}_0 = +\infty$, then also $\hat{\mathcal{E}}_0^g(\xi) = +\infty$ and there is nothing to prove. Suppose now that $\hat{Y}_0 < \infty$. This implies that $\mathcal{A}(\xi^n, g) \neq \emptyset$, for all $n \in \mathbb{N}$. Since $\xi^n \geq \eta$, Proposition 1.4 yields $(\xi^n) \subset L^1$ and $(\mathcal{E}^g(\xi^n))$ is a well-defined increasing sequence of càdlàg supermartingales. We define $Y_t = \lim_n \mathcal{E}_t^g(\xi^n)$, for all $t \in [0,T]$. Note that $Y_0 = \hat{Y}_0$. We show that Y is a càdlàg supermartingale.

To this end, note that the sequence $\left(\mathcal{E}^g(\xi^n) - \mathcal{E}^g(\xi^1)\right)$ is positive and increases to $Y - \mathcal{E}^g(\xi^1)$. Therefore monotone convergence yields

$$0 \leq E[Y_t - \mathcal{E}_t^g(\xi^1)] = \lim_n E[\mathcal{E}_t^g(\xi^n) - \mathcal{E}_t^g(\xi^1)].$$

The supermartingale property of $\mathcal{E}^g(\xi^n)$ implies that $E[\mathcal{E}^g_t(\xi^n)] \leq \mathcal{E}^g_0(\xi^n) \leq Y_0$. Furthermore, $E[\xi^1] \leq E[\mathcal{E}^g_t(\xi^1)] \leq Y_0$ and thus

$$0 \leq E[Y_t - \mathcal{E}^g_t(\xi^1)] \leq -E[\xi^1] + Y_0 < +\infty.$$

From $\mathcal{E}^g_t(\xi^1) \in L^1$, we deduce that $Y_t \in L^1$. Since $\xi = Y_T$, this implies in particular that $\xi \in L^1$. The supermartingale property follows by a similar argument. Moreover, [Dellacherie and Meyer, 1982, Theorem VI.18] implies that Y is indistinguishable from a càdlàg process. Hence, Y is a càdlàg supermartingale.

Theorem 1.7 provides a sequence of optimal controls (Z^n) such that $(\mathcal{E}^g(\xi^n), Z^n) \in \mathcal{A}(\xi^n, g)$, for all $n \in \mathbb{N}$. Now we apply the procedure introduced in the proof of Theorem 1.7 and obtain a candidate control process $\hat{Z}$. The only notable difference in the proof, except for the fact that Y is already càdlàg, is that, here, the sequence $(\mathcal{E}^g(\xi^n))$ is increasing instead of decreasing. Thus, the càdlàg supermartingales Y and $\mathcal{E}^g(\xi^1)$ serve as upper and lower bound, respectively. Consequently, we replace Y^1 by Y and $E[\xi^- \mid \mathcal{F}_\cdot]$ by $\mathcal{E}^g(\xi^1)$ in the key Inequality (1.15). The verification follows exactly the same argumentation as in the proof of Theorem 1.7 for both monotonicity Assumptions (Inc) and (Dec). Finally, to get the admissibility of $\hat{Z}$ we denote with $(\tilde{\xi}^n)$ the sequence of convex combinations of (ξ^n) corresponding to $(\tilde{Z}^n)$. Monotonicity of the sequence (ξ^n) implies $\xi^1 \leq \tilde{\xi}^n \leq \xi$, for all $n \in \mathbb{N}$. We may and do switch to a subsequence such that $(\tilde{\xi}^n)$ is increasing as well. Now, fix an arbitrary $t \in [0,T]$. Dominated convergence implies the L^1-convergence $\lim_n E[\tilde{\xi}^n \mid \mathcal{F}_t] = E[\xi \mid \mathcal{F}_t]$. Hence, we may select a subsequence such that we have P-almost sure convergence. Similar to (1.34) this implies

$$Y_0 - \int_0^t g_u(Y_u, \hat{Z}_u) du + \int_0^t \hat{Z}_u dW_u \geq \limsup_n E\left[\tilde{\xi}^n \,\middle|\, \mathcal{F}_t\right] = E\left[\xi \,\middle|\, \mathcal{F}_t\right].$$

As before, this entails that $(Y, \hat{Z}) \in \mathcal{A}(\xi, g)$. Hence, from $\mathcal{A}(\xi, g) \neq \emptyset$ and $\xi^- \in L^1$ we derive by Theorem 1.7 that there exists a control process Z such that $(\mathcal{E}^g(\xi), Z) \in \mathcal{A}(\xi, g)$. In particular this yields $Y_0 = \mathcal{E}^g_0(\xi)$, that is $\lim_n \mathcal{E}^g_0(\xi^n) = \mathcal{E}^g_0(\xi)$, since otherwise $\mathcal{E}^g_0(\xi)$ were not optimal.

Fatou's lemma: The result follows by applying monotone convergence. Indeed, denote by ζ^n the random variables $\zeta^n = \inf_{k \geq n} \xi^k$. Then from $\liminf_n \xi^n = \lim_n \zeta^n$, $\zeta^n \geq \eta$, $\zeta^n \leq \xi^n$, for all $n \in \mathbb{N}$, and monotone convergence follows

$$\hat{\mathcal{E}}^g_0(\liminf_n \xi^n) = \hat{\mathcal{E}}^g_0(\lim_n \zeta^n) = \lim_n \hat{\mathcal{E}}^g_0(\zeta^n) \leq \liminf_n \hat{\mathcal{E}}^g_0(\xi^n). \qquad \square$$

Remark 1.12. An inspection of the proof of Theorem 1.11 shows that under the assumptions implying monotone convergence, if $\lim_n \hat{\mathcal{E}}^g_0(\xi^n) < +\infty$, then $\mathcal{A}(\xi, g) \neq \emptyset$ and $\mathcal{E}^g_t(\xi^n)$ converges P-almost surely to $\mathcal{E}^g_t(\xi)$, for all $t \in [0,T]$.

Similarly, given a sequence $((Y^n, Z^n)) \subset \mathcal{A}(\xi, g)$ such that (Y^n) is increasing and $\lim_n Y^n_0 < \infty$, then there exists a control process $Z \in \mathcal{L}$ such that $(Y, Z) \in \mathcal{A}(\xi, g)$, where Y_t is the P-almost sure limit of (Y^n_t), for all $t \in [0,T]$. ♦

A consequence of the preceding theorem is the following result on L^1-lower semicontinuity.

Theorem 1.13. *Let g be a generator fulfilling* (POS), (CON) *and either* (INC) *or* (DEC). *Then $\hat{\mathcal{E}}_0^g$ is L^1-lower semicontinuous.*

Proof. Let (ξ^n) be a sequence of terminal conditions, which converges in L^1 to a random variable ξ. Suppose that there exists a subsequence $(\tilde{\xi}^n) \subset (\xi^n)$ such that $(\hat{\mathcal{E}}_0^g(\tilde{\xi}^n))$ converges to some real $a < \hat{\mathcal{E}}_0^g(\xi)$. We can assume, up to another fast subsequence, that $\|\tilde{\xi}^n - \xi\|_{L^1} \leq 2^{-n}$, for all $n \in \mathbb{N}$. Consider now the sequence (ζ^n), with ζ^n given by

$$\zeta^n = \xi - \sum_{k \geq n} (\tilde{\xi}^k - \xi)^-.$$

Clearly, $\zeta^n \in L^1$ and $\zeta^n \leq \zeta^{n+1} \leq \cdots \leq \xi$. Moreover, (ζ^n) converges in L^1 to ξ, and, since it is increasing, it converges also P-almost surely. Thus, from Theorem 1.11, we get $\lim_n \hat{\mathcal{E}}_0^g(\zeta^n) = \hat{\mathcal{E}}_0^g(\xi)$. Now, $\zeta^n \leq \xi - (\tilde{\xi}^n - \xi)^- + (\tilde{\xi}^n - \xi)^+ \leq \xi^n$ and monotony of the functional $\hat{\mathcal{E}}_0^g$ imply $a = \lim_n \hat{\mathcal{E}}_0^g(\tilde{\xi}^n) \geq \lim_n \hat{\mathcal{E}}_0^g(\zeta^n) = \hat{\mathcal{E}}_0^g(\xi)$, which is a contradiction. Hence, $\liminf_n \hat{\mathcal{E}}_0^g(\xi^n) \geq \hat{\mathcal{E}}_0^g(\xi)$. □

The preceding results allows to derive a dual representation, by means of the Fenchel-Moreau theorem, of the functional $\hat{\mathcal{E}}^g(\cdot)$ at time zero.

Corollary 1.14. *Let g be a generator fulfilling* (POS) *and either* (INC) *or* (DEC). *Assume that g is jointly convex in y and z. Then, either $\hat{\mathcal{E}}_0^g \equiv +\infty$ or*

$$\hat{\mathcal{E}}_0^g(\xi) = \mathcal{E}_0^g(\xi) = \inf_{\nu \in L_+^\infty} \left\{ E[\nu\xi] - \left(\hat{\mathcal{E}}_0^g\right)^*(\nu) \right\}, \quad \xi \in L^1, \tag{1.37}$$

for the conjugate $(\hat{\mathcal{E}}_0^g)^*(\nu) = \inf_{\xi \in L^1} \left\{ E[\nu\xi] - \hat{\mathcal{E}}_0^g(\xi) \right\}$, where $\nu \in L^\infty$.

Proof. Since $\hat{\mathcal{E}}_0^g > +\infty$ on L^1, either $\hat{\mathcal{E}}_0^g \equiv +\infty$ or $\hat{\mathcal{E}}_0^g$ is proper. In the latter case, in view of Proposition 1.2 and Theorem 1.13, the function $\hat{\mathcal{E}}_0^g$ is convex and $\sigma(L^1, L^\infty)$-lower semicontinuous on L^1. Hence, the Fenchel Moreau theorem yields the dual representation (1.37). That the domain of $(\hat{\mathcal{E}}_0^g)^*$ is concentrated on L_+^∞ follows from the monotonicity of $\hat{\mathcal{E}}_0^g$, see Proposition 1.2. □

Remark 1.15. Notice that, if the generator in Corollary 1.14 does not depend on y, then by Item 5 of Proposition 1.2 the operator $\hat{\mathcal{E}}_0^g(\cdot)$ is translation invariant. Therefore, it is a lower semicontinuous, convex risk measure and the Representation (1.37) corresponds to the robust representation of lower semicontinuous, convex risk measures; see Föllmer and Schied [2004]. ♦

Under additional integrability assumptions on the terminal condition we may also formulate stability results for supersolutions in the set $\mathcal{A}^1(\xi, g)$ introduced in (1.35).

Theorem 1.16. *Suppose that the generator g fulfills* (POS), (CON) *and either* (DEC) *or* (INC). *Let (ξ^n) be a sequence in L^0, such that $\xi^n \geq \eta$, for all $n \in \mathbb{N}$, where $(E[\eta \mid \mathcal{F}_\cdot])_T^* \in L^1$.*

- *Suppose* (ξ^n) *is increasing* P*-almost surely to* $\xi \in L^0$*, where* $(E[\xi^- \mid \mathcal{F}.])^*_T \in L^1$ *and* $\mathcal{A}^1(\xi, g) \neq \emptyset$*. Then* $\mathcal{E}^g_t(\xi) = \lim_n \mathcal{E}^g_t(\xi^n)$*,* P*-almost surely, for all* $t \in [0, T]$*.*

- *Suppose* $(E[(\liminf_n \xi^n)^- \mid \mathcal{F}.])^*_T \in L^1$ *and* $\mathcal{A}^1(\liminf_n \xi^n, g) \neq \emptyset$*.*
  Then $\mathcal{E}^g_t(\liminf_n \xi^n) \leq \liminf_n \mathcal{E}^g_t(\xi^n)$*,* P*-almost surely, for all* $t \in [0, T]$*.*

We omit the proof of the preceding theorem, as it is a simple adaptation of the proofs of Theorems 1.9 and 1.11. Note that Theorem 1.16 is a weaker version of Theorem 1.11. Indeed, here, given a sequence (ξ^n) increasing to ξ, we need to assume that $\mathcal{A}^1(\xi, g)$ is not empty. The underlying reason being the lack of knowledge whether the limit process Y, defined in the proof of Theorem 1.11, fulfills $Y^*_T \in L^1$.

The theorem above allows to state the following result on $\|\cdot\|_{\mathcal{L}^1}$-lower semicontinuity of $\hat{\mathcal{E}}^g$. Its proof is virtually the same as the proof of Theorem 1.13.

Theorem 1.17. *Suppose that the generator* g *fulfills* (Pos)*,* (Con) *and either* (Dec) *or* (Inc)*. Then* $\xi \mapsto \hat{\mathcal{E}}^g_0(\xi)$ *is* $\|\cdot\|_{\mathcal{L}^1}$*-lower semicontinuous on its domain, that is on*

$$\left\{\xi \in L^0 : (E[\xi^- \mid \mathcal{F}.])^*_T \in L^1 \text{ and } \mathcal{A}^1(\xi, g) \neq \emptyset\right\}. \tag{1.38}$$

We conclude this section with a theorem on monotone stability with respect to the generator.

Theorem 1.18. *Let* $\xi \in L^0$ *be a terminal condition, such that* $\xi^- \in L^1$*, and let* (g^n) *be an increasing sequence of generators, which converge pointwise to a generator* g*. Suppose that each generator fulfills* (Pos)*,* (Con) *and either* (Inc) *or* (Dec)*. Then* $\lim_n \hat{\mathcal{E}}^{g^n}_0(\xi) = \hat{\mathcal{E}}^g_0(\xi)$*. If, in addition,* $\lim_n \hat{\mathcal{E}}^{g^n}_0(\xi) < \infty$*, then* $\mathcal{A}(\xi, g) \neq \emptyset$ *and* $\mathcal{E}^{g^n}_t(\xi)$ *converges* P*-almost surely to* $\mathcal{E}^g_t(\xi)$*, for all* $t \in [0, T]$*.*

Proof. Note that from Proposition 1.2, we have $\hat{\mathcal{E}}^{g^n}(\xi) \leq \hat{\mathcal{E}}^{g^{n+1}}(\xi) \leq \cdots \leq \hat{\mathcal{E}}^g(\xi)$. Hence, we may set $\hat{Y}_0 = \lim_n \hat{\mathcal{E}}^{g^n}_0(\xi)$. If $\hat{Y}_0 = \infty$, then also $\hat{\mathcal{E}}^g_0(\xi) = \infty$ and we are done. Suppose that $\hat{Y}_0 < \infty$. By the same arguments as in the proof of Theorem 1.11, we constructa a càdlàg supermartingale Y. With the same procedure as in Theorem 1.11, we construct the candidate $\hat{Z}$. It remains to show $(Y, \hat{Z}) \in \mathcal{A}(\xi, g)$. However, this can be done similarly as in the proof of Theorem 1.7. We only show how to obtain the analogue of (1.33). Note first that the pointwise convergence of the generators implies that $(g^k(Y, \hat{Z}))$ converges $P \otimes dt$-almost surely to $g(Y, \hat{Z})$. Hence, Fatou's lemma yields

$$Y_s - \int_s^t g_u(Y_u, \hat{Z}_u)du + \int_s^t \hat{Z}_u dW_u \geq \limsup_k \left(Y_s - \int_s^t g^k_u(Y_u, \hat{Z}_u)du + \int_s^t \hat{Z}_u dW_u \right). \tag{1.39}$$

As in the previous proof, we use the expression in the bracket on the right hand side to obtain

$$Y_s - \int_s^t g_u^k(Y_u, \hat{Z}_u)du + \int_s^t \hat{Z}_u dW_u \geq \limsup_n \sum_{i=n}^{M^{(n)}} \lambda_i^{(n)} \left(Y_s^i - \int_s^t g_u^k(Y_u^i, Z_u^i)du + \int_s^t Z_u^i dW_u \right).$$

Since on the right hand side we consider the lim sup with respect to n and k being fixed for the moment, we may assume $k \leq n$, which entails by monotonicity of the sequence of generators

$$Y_s - \int_s^t g_u^k(Y_u, \hat{Z}_u)du + \int_s^t \hat{Z}_u dW_u \geq \limsup_n \sum_{i=n}^{M^{(n)}} \lambda_i^{(n)} \left(Y_s^i - \int_s^t g_u^i(Y_u^i, Z_u^i)du + \int_s^t Z_u^i dW_u \right).$$

From here, we obtain as before $Y_s - \int_s^t g_u^k(Y_u, \hat{Z}_u)du + \int_s^t \hat{Z}_u dW_u \geq Y_t$, where the right hand side does not depend on k anymore. Combined with (1.39), this yields the analogue of (1.33). □

Remark 1.19. Similar to Theorem 1.11 one may formulate extensions of Theorem 1.18, allowing not only for monotone increasing approximations of the limit generator g. Consider for example a sequence of generators (g^n) and a generator g, that neither depend on y nor, for simplicity, on (ω, t). If in addition to $\lim_n g^n(z) \to g(z)$ we have $\lim_n h^n(z) \to g(z)$, where h^n is defined by $h^n = \operatorname{conv}\left\{g, g^n, g^{n+1}, \cdots\right\}$, then by Proposition 1.2 and Theorem 1.18, $\liminf_n \hat{\mathcal{E}}_0^{g^n}(\xi) \geq \liminf_n \hat{\mathcal{E}}_0^{h^n}(\xi) = \hat{\mathcal{E}}_0^{g}(\xi)$. To give precise conditions in a general setting under which the convergence of g^n to g implies convergence of h^n to g is not in line with our main objective and consequently this question is left for future investigations. ♦

1.3.3. Non positive generators

In this section we extend our results to generators that are not necessarily positive. This is important with regards to applications in mathematical finance, where the generators are quite often of the linear-quadratic type. It turns out that we can extend the scope of our theorems to cover precisely some of these situations; see Section 1.3.4 for such an example. Using some measure change, the positivity assumption on the generator g can be relaxed to a linear bound below. This leads to optimal solutions under P, where the admissibility is required with respect to the related equivalent probability measure. More precisely, we say in the following that a generator g is

(LB) linearly bounded from below, if there exist adapted measurable $\mathbb{R}^{1\times d}$ and $\mathbb{R}$-valued processes a and b, respectively, such that $g(y,z) \geq az^\top + b$, for all $y, z \in \mathbb{R}\times\mathbb{R}^{1\times d}$. Furthermore, $\int_0^t b_s ds \in L^1(P^a)$, for all $t\in[0,T]$ and

$$\frac{dP^a}{dP} = \mathcal{E}\left(\int adW\right)_T,$$

defines an equivalent probability measure P^a.

Example 1.20. For instance, given a generator g, assume that there exists a generator $\hat{g}$ independent of y fulfilling (CON) and such that $g \geq \hat{g}$. Then, there exists an $\mathbb{R}^{1\times d}$-valued adapted measurable process a such that $g(y,z) \geq az^\top - \hat{g}^*(a)$, for all $y, z \in \mathbb{R}\times\mathbb{R}^{1\times d}$, where $\hat{g}^*$ denotes the convex conjugate of g. ♢

In the following, we say that Z is a-admissible, if $\int ZdW^a$ is a P^a-supermartingale, where $W^a = (W^1 - \int a^1 ds, \cdots, W^d - \int a^d ds)^\top$ is the respective Brownian motion under P^a. We are interested in the sets

$$\mathcal{A}^a(\xi,g) = \{(Y,Z)\in\mathcal{S}\times\mathcal{L} : Z \text{ is } a\text{-admissible and (1.2) holds}\}, \tag{1.40}$$

and define the random process

$$\hat{\mathcal{E}}^{g,a}_t(\xi) = \operatorname{ess\,inf}\left\{Y_t\in L^0(\mathcal{F}_t) : (Y,Z)\in\mathcal{A}^a(\xi,g)\right\}, \quad t\in[0,T]. \tag{1.41}$$

The analogue of Theorem 1.7 is given as follows

Theorem 1.21. *Let g be a generator fulfilling* (LB), (CON) *and either* (INC) *or* (DEC) *and $\xi\in L^0$ be a terminal condition, such that $\xi^-\in L^1(P^a)$. If $\mathcal{A}^a(\xi,g)\neq\emptyset$, then there exists a unique minimal supersolution $(\hat{Y},\hat{Z})\in\mathcal{A}^a(\xi,g)$. Moreover, $\mathcal{E}^g(\xi)$ is the value process of the minimal supersolution, that is $(\mathcal{E}^g(\xi),\hat{Z})\in\mathcal{A}^a(\xi,g)$.*

The analogues of Theorem 1.11 and Theorem 1.13 read as follows.

Theorem 1.22. *Suppose that the generator g fulfills* (LB), (CON) *and either* (INC) *or* (DEC). *Let (ξ^n) be a sequence in L^0, such that $\xi^n\geq\eta$, for all $n\in\mathbb{N}$, where $\eta\in L^1(P^a)$.*

- *Monotone convergence: If (ξ^n) is increasing P-almost surely to $\xi\in L^0$, then $\hat{\mathcal{E}}^{g,a}_0(\xi) = \lim_n \hat{\mathcal{E}}^{g,a}_0(\xi^n)$.*
- *Fatou's lemma: $\hat{\mathcal{E}}^{g,a}_0(\liminf_n \xi^n) \leq \liminf_n \hat{\mathcal{E}}^{g,a}_0(\xi^n)$.*

In particular, $\hat{\mathcal{E}}^{g,a}_0$ is $L^1(P^a)$-lower semicontinuous.

We only prove the first theorem.

Proof (of Theorem 1.21). In the setting of Section 1.3.1, given a positive generator $\bar{g}$ and a random variable ζ, let us denote by $\mathcal{A}(\zeta,\bar{g},W^a)$ the set defined in (1.4) to indicate the dependence of this set on the Brownian motion W^a and the respective probability measure P^a. Let us now define the generator $\bar{g}$ as

$$\bar{g}(y,z) = g\left(y + \int_0^{\cdot} b_s ds, z\right) - az^\top - b, \quad \text{for all } (y,z) \in \mathbb{R} \times \mathbb{R}^{1\times d}. \tag{1.42}$$

By Assumption (LB), this generator fulfills (POS), (CON) and either (INC) or (DEC). Since $\int Z dW^a$ is a P^a-supermartingale, a simple inspection shows that the affine transformation $\bar{Y} = Y - \int b ds$ and $\bar{Z} = Z$ yields a one-to-one relation between $\mathcal{A}^a(\xi, g)$ and $\mathcal{A}(\xi - \int_0^T b_s ds, \bar{g}, W^a)$. Hence, the assumptions of Theorem 1.7 are fulfilled for $\bar{g}$ and $\mathcal{A}(\xi - \int_0^T b_s ds, \bar{g}, W^a)$, and thus its application ends the proof. □

Remark 1.23. Note that if $(E^a[(\xi - \int_0^T b_s ds)^- \mid \mathcal{F}_\cdot])_T^* \in L^1(P^a)$, then Theorem 1.9 applies in the same way, that is, under the assumptions of Theorem 1.21, if

$$\mathcal{A}^{1,a}(\xi, g) := \left\{(Y,Z) \in \mathcal{A}^a(\xi, g) : Z \in \mathcal{L}^1(P^a)\right\} \neq \emptyset,$$

then $\mathcal{E}^{g,a}(\xi)$ is the value process of the minimal supersolution with unique control process $Z \in \mathcal{L}^1(P^a)$. ⧫

1.3.4. Expected exponential utility maximization

In this section, we illustrate our results, in particular our set based comparison principle and the existence theorem, by maximizing exponential expected utility. For an introduction to this well-known problem and for similar statements, see for instance Delbaen et al. [2002], Hu et al. [2005], Mania and Schweizer [2005], and the references therein.

Consider a financial market where the discounted stock prices are modelled by a n-dimensional process S satisfying the differential equation

$$dS_u^i = S_u^i\left(\mu_u^i du + \sigma_u^i dW_u\right), \quad S_0^i > 0, \quad i = 1, \ldots, n, \tag{1.43}$$

where $n \leq d$, and the drift $\mu = (\mu^1, \ldots, \mu^n)^\top$ and the volatility row-vectors $\sigma^i = (\sigma^{i1}, \ldots, \sigma^{id})$ are progressive processes. We suppose that $\sigma\sigma^\top$ is invertible, $P \otimes dt$-almost surely. Hence, without loss of generality, we may assume that $\sigma^{ij} = 0$, for all $j = n+1, \ldots, d$, and that the $n\times n$-matrix $\tilde{\sigma}^{ij} = \sigma^{ij}$, $i,j = 1, \ldots, n$, is invertible $P\otimes dt$-almost surely. By use of the d-dimensional market price of risk vector $\theta = (\tilde{\sigma}^{-1}\mu, 0, \ldots, 0)^\top$, the dynamics in (1.43) are equivalent to $dS_u^i = S_u^i \sigma_u^i(\theta_u du + dW_u)$, for $i = 1, \ldots, n$. We further assume that[1] $\int \theta^\top dW \in BMO(P)$. Therefore, $\mathcal{E}\left(-\int \theta^\top dW\right)_T$ defines an equivalent probability measure $P^{-\theta}$. According to Girsanov's change of measure theorem, the discounted price process S is then a local martingale under $P^{-\theta}$. We

[1] For the definition and important results concerning the space BMO, we refer to Kazamaki [1994].

consider the problem

$$U_t^\vartheta = E\left[\exp\left(-\xi - \int_0^T \vartheta_u dS_u\right) \Big| \mathcal{F}_t\right] \to \min, \tag{1.44}$$

where $\xi \in L^0(\mathcal{F}_T)$ is a random endowment and ϑ an $\mathbb{R}^{1\times n}$-valued admissible strategy. Here, the optimization takes place over the set Θ of those strategies ϑ such that ϑ is progressive, $\int \vartheta dS$ is a $P^{-\theta}$-supermartingale and $\exp(-\int_0^\tau \vartheta dS))_{\tau\in\mathcal{T}}$ is P-uniformly integrable.

In principle, the following proposition is well-known, see the references mentioned above. However, our approach relies exclusively on the theory of supersolutions developed in the previous sections and extends the method of pointwise optimization of generators used in Horst et al. [2010] to our setting of an incomplete market.

Proposition 1.24. *Suppose that $\int \theta^\top dW \in BMO(P)$ and $\xi \in L^\infty$. Then, ϑ^* defined by the following componentwise multiplication between ϑ^* and S,*

$$\vartheta_t^* S_t = (Z_t + \theta_t^\top)\begin{pmatrix}\tilde{\sigma}_t^{-1}\\ 0\end{pmatrix} = (Z_t^1,\dots,Z_t^n)\tilde{\sigma}_t^{-1} + (\theta_t^1,\dots,\theta_t^n)\tilde{\sigma}_t^{-1}, \tag{1.45}$$

is the optimal solution in Θ of (1.44). Here, $Z \in \mathcal{L}^1(P^{-\theta})$ is the control process corresponding to the minimal supersolution of the BSDE

$$Y_s - \int_s^t \left[\frac{1}{2}\left(\sum_{k=n+1}^d (Z_u^k)^2 - \sum_{k=1}^n (\theta_u^k)^2\right) - \sum_{k=1}^n \theta_u^k Z_u^k\right] du + \int_s^t Z_u dW_u \geq Y_t$$
$$\text{and} \quad Y_T \geq -\xi. \tag{1.46}$$

Proof. Let $\vartheta \in \Theta^\infty$, the set of those strategies in Θ such that $\int \vartheta dS$ is uniformly bounded. The general result follows by stopping any strategy ϑ as soon as $|\int \vartheta dS|$ is above $n \in \mathbb{N}$ and using the uniform integrability to show that $U_t^{\vartheta^n}$ converges P-almost surely to U_t^ϑ, for all $t \in [0,T]$, where (ϑ^n) is the stopped strategy. In view of Itô's lemma, the certainty equivalent $Y_t^\vartheta = \log(U_t^\vartheta) + \int_0^t \vartheta_u dS_u$ satisfies the BSDE

$$Y_s^\vartheta - \int_s^t g_u\left(\vartheta_u, Z_u^\vartheta\right) du + \int_s^t Z_u^\vartheta dW_u = Y_t^\vartheta, \quad Y_T^\vartheta = -\xi, \tag{1.47}$$

where the control process is given by $Z_t^\vartheta = V_t^\vartheta / U_t^\vartheta + \vartheta_t S_t \sigma_t$, with V^ϑ coming from the martingale representation of U^ϑ. The generator is given by

$$g_u(\vartheta, z) = \frac{1}{2}\left(z - \vartheta_u S_u \sigma_u\right)^2 - \vartheta_u S_u \mu_u,$$

for all $z \in \mathbb{R}^{1\times d}$, and, from our assumption on θ, it follows that $\int Z^\vartheta dW^{-\theta}$ is a $P^{-\theta}$-

supermartingale. This implies that $(Y^\vartheta, Z^\vartheta) \in \mathcal{A}^{-\theta}(-\xi, g(\vartheta,\cdot))$. Pointwise optimization of $\vartheta \mapsto \frac{1}{2}(z - \vartheta S_u \sigma_u)^2 - \vartheta S_u \mu_u$, for $\vartheta \in \mathbb{R}^{1\times n}$, yields a pointwise minimum $\vartheta^*(z)$ such that

$$\vartheta^*_u(z) S_u = (z + \theta_t^\top) \begin{pmatrix} \tilde{\sigma}_t^{-1} \\ 0 \end{pmatrix}.$$

Plugged into the generator $g_u(\vartheta, z)$, it yields the optimized generator

$$g^*_u(z) = \frac{1}{2}\left(\sum_{k=n+1}^{d}(z^k)^2 - \sum_{k=1}^{n}(\theta_u^k)^2\right) - \sum_{k=1}^{n}\theta_u^k z^k.$$

Now we use our comparison result, see Proposition 1.2, to obtain

$$\mathcal{A}^{-\theta}(-\xi, g(\vartheta,\cdot)) \subset \mathcal{A}^{-\theta}(-\xi, g^*), \qquad \text{for all } \vartheta \in \Theta^\infty,$$

and

$$\hat{\mathcal{E}}^{g^*,-\theta}(-\xi) \leq Y^\vartheta,$$

for all $\vartheta \in \Theta^\infty$. Next we want to use existence and uniqueness of the minimal supersolution given by Theorem 1.21 to ensure that $\hat{\mathcal{E}}^{g^*,-\theta}(-\xi)$ has indeed a modification $\mathcal{E}^{g^*,-\theta}(-\xi)$ which is the value process of the minimal supersolution and to obtain the corresponding control process. To that end note the generator g^* fulfills

$$g^*_u(z) \geq -\frac{1}{2}\sum_{k=1}^{n}(\theta_u^k)^2 - \theta^\top z^\top$$

and that this is a valid lower bound in the sense of Assumption (LB) in Section 1.3.3. From our assumptions on θ and ξ follows

$$(E^{-\theta}[|\xi| + \frac{1}{2}\int_0^T \sum_{k=1}^{n}(\theta_u^k)^2 du \mid \mathcal{F}_\cdot])^*_T \in L^1(P^{-\theta}).$$

This and ξ bounded below imply that

$$(\|\xi^-\|_\infty - \frac{1}{2}\int \sum_{k=1}^{n}(\theta^k)^2 du, 0) \in \mathcal{A}^{1,-\theta}(-\xi, g^*),$$

that is $\mathcal{A}^{1,-\theta}(-\xi, g^*) \neq \emptyset$. Hence, all the assumptions of Theorem 1.21 are fulfilled, which, together with Remark 1.23, yields that $Y^* := \mathcal{E}^{g^*,-\theta}(-\xi)$ is the value process of the minimal supersolution of (1.46) with control process $Z^* \in \mathcal{L}^1(P^{-\theta})$. Defining $\vartheta^* := \vartheta^*(Z^*)$, yields

$$(Y^*, Z^*) = (Y^{\vartheta^*}, Z^*) \in \mathcal{A}^{1,-\theta}(-\xi, g(\vartheta^*,\cdot)).$$

We finally have to show that $\vartheta^* \in \Theta$ and that $U_0^{\vartheta^*}$ is minimal. From $\int \theta^\top dW^{-\theta} \in BMO(P^{-\theta})$ it follows that $\int \vartheta^* dS = \sum_{k=1}^n \int (Z^{*k} + \theta^k) dW^{-\theta,k} \in \mathcal{L}^1(P^{-\theta})$. Thus,

$\int \vartheta^* dS$ is a $P^{-\theta}$-(super)martingale. Furthermore, by (1.46), for all stopping times $\tau \in \mathcal{T}$, holds

$$E^{-\theta}\left[\int_\tau^T \sum_{k=n+1}^d (Z^{*k})^2 du \mid \mathcal{F}_\tau\right] \leq 2Y^*_\tau + 2E^{-\theta}[\xi \mid \mathcal{F}_\tau] + E^{-\theta}\left[\int_\tau^T \sum_{k=n+1}^d (\theta^k)^2 du \mid \mathcal{F}_\tau\right].$$

From $Y^* \leq \|\xi^-\|_\infty$ and $\xi \leq \|\xi^+\|_\infty$, it follows

$$\left(\int Z^{*n+1} dW^{-\theta,n+1}, \cdots, \int Z^{*d} dW^{-\theta,d}\right) \in BMO(P^{-\theta}).$$

Thus, for $\tilde{V}^* := Z^* - \vartheta^* S\sigma$, it holds

$$\int \tilde{V}^* dW^{-\theta} = -\sum_{k=1}^n \int \theta^k dW^{-\theta,k} + \sum_{k=n+1}^d \int Z^{*k} dW^{-\theta,k} \in BMO(P^{-\theta}),$$

and therefore, $\int \tilde{V}^* dW \in BMO(P)$. Since $\int \vartheta^* dS = \int \vartheta^* S\sigma dW^{-\theta}$ and $\vartheta^* S\sigma = Z^* + \tilde{V}^*$, the admissibility of Z^* yields that $\int \vartheta^* dS$ is a $P^{-\theta}$-supermartingale. The process $C^{\vartheta^*} = Y^* - \int \vartheta^* dS$ satisfies

$$C^{\vartheta^*}_s - \int_s^t \frac{1}{2}(\tilde{V}^*_u)^2 du + \int_s^t \tilde{V}^*_u dW_u \geq C^{\vartheta^*}_t, \quad C^{\vartheta^*}_T = -\xi - \int_0^T \vartheta^*_u dS_u. \tag{1.48}$$

Consequently,

$$\exp\left(-\int_0^t \vartheta^* dS\right) = \exp\left(C^{\vartheta^*}_t - Y^*_t\right) \leq \mathcal{E}\left(\int \tilde{V}^* dW\right)_t \exp\left(C^{\vartheta^*}_0 - Y^*_t\right).$$

Hence, from $\int \tilde{V}^* dW \in BMO(P)$ and $Y^* \geq -\|\xi\|_\infty$, it follows that the family $(\exp(-\int_0^\tau \vartheta^* dS))_{\tau\in\mathcal{T}}$ is P-uniformly integrable. Thus, $\vartheta^* \in \Theta$. As for the optimality of ϑ^*, note that, for $\vartheta \in \Theta^\infty$, holds

$$\begin{aligned} E\left[\exp\left(-\xi - \int_0^T \vartheta^*_u dS_u\right) \;\middle|\; \mathcal{F}_t\right] &= E\left[\exp\left(C^{\vartheta^*}_T\right) \;\middle|\; \mathcal{F}_t\right] \\ &\leq \exp\left(C^{\vartheta^*}_t\right) = \exp\left(Y^*_t - \int_0^t \vartheta^*_u dS_u\right) \leq \exp\left(Y^\vartheta_t - \int_0^t \vartheta^*_u dS_u\right), \end{aligned}$$

which implies

$$E\left[\exp\left(-\xi - \int_t^T \vartheta^*_u dS_u\right) \;\middle|\; \mathcal{F}_t\right] \leq E\left[\exp\left(-\xi - \int_t^T \vartheta_u dS_u\right) \;\middle|\; \mathcal{F}_t\right]. \qquad \square$$

1.4. Helly's theorem

In principle the following result is well-known, see for example [Campi and Schachermayer, 2006, De Vallière et al., 2009, Kupper and Schachermayer, 2009] for similar statements. We give a proof for sake of completeness.

Lemma 1.25. *Let (A^n) be a sequence of increasing positive processes such that the sequence (A^n_T) is bounded in L^1. Then, there is a subsequence $(\tilde{A}^{n_k})$ in the asymptotic convex hull of (A^n) and an increasing positive integrable process $\tilde{A}$ such that*

$$\lim_{k\to\infty} \tilde{A}^{n_k}_t = \tilde{A}_t, \quad \text{for all } t \in [0,T], \quad P\text{-almost surely.} \tag{1.49}$$

Proof. Let (t_j) be a sequence running through $\mathcal{I} = ([0,T]\cap\mathbb{Q})\cup\{T\}$. Since $(A^n_{t_1})$ is an L^1-bounded sequence of positive random variables, due to [Delbaen and Schachermayer, 1994, Lemma A1.1], there exists a subsequence $(\tilde{A}^{1,k})$ in the asymptotic convex hull of (A^n) and a random variable $\tilde{A}_{t_1}$ such that $(\tilde{A}^{1,k}_{t_1})$ converges P-almost surely to $\tilde{A}_{t_1}$. Moreover, Fatou's lemma yields $\tilde{A}_{t_1} \in L^1$. Let $(\tilde{A}^{2,k})$ be a subsequence in the asymptotic convex hull of $(\tilde{A}^{1,k})$ such that $(\tilde{A}^{2,k}_{t_2})$ converges P-almost surely to $\tilde{A}_{t_2} \in L^1$ and so on. Then, $\tilde{A}^{k,k}_t \to \tilde{A}_t$, on a set $\hat{\Omega} \subset \Omega$, where $P(\hat{\Omega}) = 1$. The process $\tilde{A}$ is positive, increasing and integrable on $\mathcal{I}$. Thus, we may define

$$\hat{A}_t = \lim_{r\downarrow t, r\in\mathcal{I}} \tilde{A}_r, \quad t \in [0,T), \quad \hat{A}_T = \tilde{A}_T.$$

We now show that $(\tilde{A}^{k,k})$, henceforth named $(\tilde{A}^k)$, converges P-almost surely on the continuity points of $\hat{A}$. Fix $\omega \in \hat{\Omega}$ and a continuity point $t \in [0,T)$ of $\hat{A}(\omega)$. We show that $(\tilde{A}^k_t(\omega))$ is a Cauchy sequence in $\mathbb{R}$.

Fix $\varepsilon > 0$ and set $\delta = \varepsilon/11$. Since t is a continuity point of $\hat{A}(\omega)$, we may choose $p_1, p_2 \in \mathcal{I}$ such that $p_1 < t < p_2$ and $\hat{A}_{p_2}(\omega) - \hat{A}_{p_1}(\omega) < \delta$. By definition of $\hat{A}$, we may choose $r_1, r_2 \in \mathcal{I}$ such that $p_1 < r_1 < t < p_2 < r_2$ and $|\hat{A}_{p_2}(\omega) - \tilde{A}_{r_2}(\omega)| < \delta$ and $|\hat{A}_{p_1}(\omega) - \tilde{A}_{r_1}(\omega)| < \delta$. Now choose $N \in \mathbb{N}$ such that $|\tilde{A}^m_{r_1}(\omega) - \tilde{A}^n_{r_1}(\omega)| < \delta$, for all $m, n \in \mathbb{N}$ with $m, n \geq N$, and $|\tilde{A}^j_{r_2}(\omega) - \tilde{A}_{r_2}(\omega)| < \delta$ and $|\tilde{A}_{r_1}(\omega) - \tilde{A}^j_{r_1}(\omega)| < \delta$, for $j = m, n$. We estimate

$$|\tilde{A}^m_t(\omega) - \tilde{A}^n_t(\omega)| \leq |\tilde{A}^m_t(\omega) - \tilde{A}^m_{r_1}(\omega)| + |\tilde{A}^m_{r_1}(\omega) - \tilde{A}^n_{r_1}(\omega)| + |\tilde{A}^n_{r_1}(\omega) - \tilde{A}^n_t(\omega)|.$$

For the first and the third term on the right hand side, since $\tilde{A}^m$ and $\tilde{A}^n$ are increasing, we deduce that $|\tilde{A}^m_t(\omega) - \tilde{A}^m_{r_1}(\omega)| \leq |\tilde{A}^m_{r_2}(\omega) - \tilde{A}^m_{r_1}(\omega)|$ and $|\tilde{A}^n_t(\omega) - \tilde{A}^n_{r_1}(\omega)| \leq |\tilde{A}^n_{r_2}(\omega) - \tilde{A}^n_{r_1}(\omega)|$. Moreover

$$\begin{aligned} |\tilde{A}^j_{r_2}(\omega) - \tilde{A}^j_{r_1}(\omega)| \leq{}& |\tilde{A}^j_{r_2}(\omega) - \tilde{A}_{r_2}(\omega)| + |\tilde{A}_{r_2}(\omega) - \hat{A}_{p_2}(\omega)| \\ &+ |\hat{A}_{p_2}(\omega) - \hat{A}_{p_1}(\omega)| + |\hat{A}_{p_1}(\omega) - \tilde{A}_{r_1}(\omega)| + |\tilde{A}_{r_1}(\omega) - \tilde{A}^j_{r_1}(\omega)|, \end{aligned}$$

for $j = m, n$. Combining the previous inequalities yields $|\tilde{A}^m_t(\omega) - \tilde{A}^n_t(\omega)| \leq \varepsilon$, for all $n, m \geq N$. Hence, $(\tilde{A}^k(\omega))$ converges for all continuity points $t \in [0,T)$ of $\hat{A}(\omega)$, for all

$\omega \in \hat{\Omega}$. We denote the limit with $\tilde{A}$.

It remains to see that $(\tilde{A}^k)$ also converges for the discontinuity points of $\hat{A}$. To this end, note that $\hat{A}$ is càdlàg and adapted to our filtration which fulfills the usual conditions. By a well-known result, see for example [Karatzas and Shreve, 2004, Proposition 1.2.26], this implies that the jumps of $\hat{A}$ may be exhausted by a sequence of stopping times (ρ^j). Applying once more [Delbaen and Schachermayer, 1994, Lemma A1.1] iteratively on the sequences $(\tilde{A}^k_{\rho^j})_{k\in\mathbb{N}}$, $j = 1, 2, 3, \ldots$, and diagonalizing yields the result. □

Part II.

Cross Hedging

2. Futures Cross-hedging with a Stationary Spread

In the introduction we have seen that the estimated correlation between the log prices of kerosene and crude oil strongly depends on the sampling frequency. More precisely, the short-term correlations are considerably lower than the long-term correlations, pointing towards a long-term relationship with potential short-term deviations. Motivated by this observation we set up a model, with mean reverting, or asymptotic stationary spread, that allows a rigorous study of the effect of a long-term relationship on optimal cross-hedging strategies, and at the same time allows an efficient calculation of the basis risk entailed by the optimal cross-hedges.

The chapter is structured as follows. Section 2.1 introduces our model and presents some empirical evidence, while Section 2.2 briefly reviews hedging with futures contracts and derives the variance optimal hedging strategy for our model. Section 2.3 develops the implied hedge errors within our model for linear and non-linear positions and Section 2.4 compares the hedge errors between different models and (suboptimal) hedging strategies emphasizing the importance of allowing for a stationary spread. An extension of our model to account for stochastic volatility is given in Section 2.5. Section 2.6 concludes.

2.1. The Continuous-time Model with a Stationary Spread

As always in modeling real-world phenomena there is a trade-off between the accuracy of a model and its tractability. We therefore illustrate the implications of an asymptotic stationary logspread between the futures price and the price of an illiquid asset by assuming a simplified and tractable model, which is presented in Section 2.1.1. This approach allows us to derive not only optimal hedging strategies and their implied hedge errors (see Section 2.2), but also to obtain the transition density in closed form, so that the model can straightforwardly be estimated via the efficient maximum likelihood method. An empirical illustration is provided in Section 2.1.2.

2.1.1. Model Specification

Let $I = (I_t)_{t\geq 0}$ denote the price process of an illiquid asset, and suppose that an economic agent aims at hedging a position $h(I_T)$, where $h : \mathbb{R} \to \mathbb{R}$ is a measurable payoff function and $T > 0$ is a fixed time horizon. Furthermore, we assume that there exists a liquidly traded futures contract with price process $X = (X_t)_{t\geq 0}$, which evolves according to

$$dX_t = \mu X_t dt + \sigma_X X_t dW_t^{(X)}, \quad X_0 = x, \tag{2.1}$$

with volatility $\sigma_X > 0$ and constant drift rate $\mu \in \mathbb{R}$. The process $W^{(X)} = \left(W_t^{(X)}\right)_{t\geq 0}$ is a Brownian motion on a stochastic basis with probability measure P. We denote the spread of the log prices, in the following simply referred to as the *logspread*, by

$$S_t = \log(X_t) - \log(I_t).$$

Although the non-stationarity of the logspread seems to be a plausible assumption for certain asset classes, e.g. for stock prices, there also exist relevant examples for asymptotic stationary logspreads as shown in the introduction and the mentioned articles. We therefore propose to account for cointegration by first modeling the logspread as an asymptotic stationary process and then derive the implied dynamics of the illiquid asset.

More precisely, we assume that the logspread follows an Ornstein-Uhlenbeck process, that is the logspread solves the SDE

$$dS_t = \kappa(m - S_t)dt + \sigma_S\left(\rho dW_t^{(X)} + \bar{\rho} dW_t^{\perp}\right), \quad S_0 = s, \tag{2.2}$$

where $W^{\perp} = \left(W_t^{\perp}\right)_{t\geq 0}$ is a Brownian motion independent of $W^{(X)}$, $\kappa \geq 0$ is the mean reversion speed, and $\rho \in [-1, 1]$ the correlation. The logspread's volatility σ_S is assumed to be non-negative. Moreover, we define $\bar{\rho} = \sqrt{1-\rho^2}$ and use, for ease of exposition, the following short-hand notation

$$W_t^{(S)} = \rho W_t^{(X)} + \bar{\rho} W_t^{\perp}$$

for the Brownian motion driving the logspread. Note that for $\kappa \downarrow 0$ the Ornstein-Uhlenbeck process becomes more and more persistent and in the limit a (scaled and shifted) Brownian motion that is correlated with the Brownian motion of the futures price process.

The dynamics of X and S determine the dynamics of the illiquid asset price, as $I_t = X_t e^{-S_t}$, $t \geq 0$. A straightforward calculation shows that the dynamics of I satisfy

$$dI_t = I_t\left(\frac{1}{2}\sigma_S^2 - \kappa(m - S_t) + \mu - \rho\sigma_S\sigma_X\right)dt + I_t\sigma_I dW_t^{(I)},$$

where $\sigma_I = \sqrt{\sigma_X^2 - 2\rho\sigma_S\sigma_X + \sigma_S^2}$ and $W^{(I)} = \left(W_t^{(I)}\right)_{t\geq 0}$ is a Brownian motion defined by $W_t^{(I)} = \left((\sigma_X - \rho\sigma_S)W_t^{(X)} - \bar{\rho}\sigma_S W_t^{\perp}\right)/\sigma_I$, $t \geq 0$.

Note that the correlation ρ_{IX} between the Brownian motions driving I and X is given by

$$\rho_{IX} = \frac{1}{\sigma_I}\left(\sigma_X - \rho\sigma_S\right), \tag{2.3}$$

which is non-negative if and only if $\sigma_X \geq \rho\sigma_S$. For fixed parameters ρ and σ_X, we can write ρ_{IX} as a function of σ_S:

$$\rho_{IX}(\sigma_S) = \frac{\sigma_X - \rho\sigma_S}{\sqrt{\sigma_X^2 - 2\rho\sigma_S\sigma_X + \sigma_S^2}}. \tag{2.4}$$

It can be shown that $\rho_{IX}(\sigma_S)$ is strictly decreasing in σ_S, and hence invertible on $\mathbb{R}_+$.

Lemma 2.1. *Let $\sigma_X > 0$ and $\rho \in (-1, 1)$. Then the mapping $\mathbb{R}_+ \ni \sigma_S \mapsto \rho_{IX}(\sigma_S)$ is strictly decreasing. Moreover, the logspread's volatility σ_S satisfies*

$$\sigma_S = \sigma_X \frac{\sqrt{1-\rho_{IX}^2}}{\rho\sqrt{1-\rho_{IX}^2} + \rho_{IX}\sqrt{1-\rho^2}}. \tag{2.5}$$

Proof. By distinguishing the cases $\rho \geq 0$ and $\rho < 0$ one can show that $\partial\rho_{IX}/\partial\sigma_S \leq 0$, and that the partial derivative is strictly smaller than zero if $\sigma_S > 0$. Thus, ρ_{IX} is strictly decreasing in σ_S. From the definition of ρ_{IX} we have

$$\rho_{IX}^2 = \left(\sigma_X - \rho\sigma_S\right)^2 / \left(\sigma_X^2 - 2\rho\sigma_S\sigma_X + \sigma_S^2\right),$$

which leads to the quadratic equation in σ_S

$$(\rho_{IX}^2 - \rho^2)\sigma_S^2 + 2\rho\sigma_X(1-\rho_{IX}^2)\sigma_S - \sigma_X^2(1-\rho_{IX}^2) = 0. \tag{2.6}$$

If $\rho \neq \rho_{IX}$, then Equation (2.6) has two solutions, namely

$$\sigma_S = \sigma_X\sqrt{1-\rho_{IX}^2}\,\frac{-\rho\sqrt{1-\rho_{IX}^2} \pm \rho_{IX}\sqrt{1-\rho^2}}{\rho_{IX}^2 - \rho^2}.$$

Since $\rho_{IX}^2 - \rho^2 = (\rho_{IX}\sqrt{1-\rho^2} + \rho\sqrt{1-\rho_{IX}^2})(\rho_{IX}\sqrt{1-\rho^2} - \rho\sqrt{1-\rho_{IX}^2})$, this further yields

$$\sigma_S = \sigma_X\sqrt{1-\rho_{IX}^2}\,\frac{1}{\rho\sqrt{1-\rho_{IX}^2} \pm \rho_{IX}\sqrt{1-\rho^2}}.$$

If $\rho_{IX} > \rho$, then only one of the roots guarantees that $\sigma_S \geq 0$, and we obtain Equation (2.5). If $\rho_{IX} = \rho$, then Equation (2.6) has a unique solution, given by $\sigma_S = \sigma_X/(2\rho)$. The inverse function of Formula (2.4) is continuous on $\rho^{-1}(\mathbb{R}_+)$. Therefore, Equation (2.5) must also hold true for $\rho_{IX} < \rho$. $\square$

Observe that $(\log X_t, S_t, \log I_t)$ is a 3-dimensional Gaussian process. Furthermore, it possesses a closed-form transition density and hence efficient maximum likelihood estimation becomes feasible. Indeed, a straightforward calculation shows that the triple $(\log X_t, S_t, \log I_t)$ satisfies, for $(X_0, S_0, I_0) = (x, s, xe^{-s})$,

$$\begin{pmatrix} \log X_t \\ S_t \\ \log I_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \log x + \left(\mu - \frac{\sigma_X^2}{2}\right)t \\ se^{-\kappa t} + m\left(1 - e^{-\kappa t}\right) \\ \log x + \left(\mu - \frac{\sigma_X^2}{2}\right)t - se^{-\kappa t} - m\left(1 - e^{-\kappa t}\right) \end{pmatrix}, \Sigma\right),$$

where, with $a_t = 1 - e^{-\kappa t}$,

$$\Sigma = \begin{pmatrix} t\sigma_X^2 & \frac{\rho\sigma_X\sigma_S}{\kappa}a_t & t\sigma_X^2 - \rho\frac{\sigma_X\sigma_S}{\kappa}a_t \\ \frac{\rho\sigma_X\sigma_S}{\kappa}a_t & \frac{\sigma_S^2}{2\kappa}a_{2t} & \frac{\rho\sigma_X\sigma_S}{\kappa}a_t - \frac{\sigma_S^2}{2\kappa}a_{2t} \\ t\sigma_X^2 - \rho\frac{\sigma_X\sigma_S}{\kappa}a_t & \frac{\rho\sigma_X\sigma_S}{\kappa}a_t - \frac{\sigma_S^2}{2\kappa}a_{2t} & t\sigma_X^2 - 2\rho\frac{\sigma_X\sigma_S}{\kappa}a_t + \frac{\sigma_S^2}{2\kappa}a_{2t} \end{pmatrix}$$

and $\mathcal{N}(m, V)$ denotes the normal distribution with mean vector m and covariance matrix V.

As we specify first the dynamics of the futures contract as a GBM and the logspread as an Ornstein-Uhlenbeck process the price of the futures contract leads the risk price. For example if the futures price is subject to a (demand or supply) shock the risk price follows and reduces the distance to the futures price. This asymmetric behavior is in line with empirical findings as there is strong evidence that futures prices lead the spot prices, e.g. see Chan [1987], Kawaller et al. [1987] and Stoll and Whaley [1990]. Note that most studies investigate the relationship between a stock index and the corresponding futures contract. However, their main argument of less frequent trading in the spot market and differences in transaction costs are also valid in our setup. They both lead to asymmetric access to information which in turn results in an asymmetric behavior of the spread. Therefore, for the applications we have in mind, it seems natural to model the futures and the spread first, and then to derive the spot dynamics endogenously. Of course, it is also possible to specify the relation in the reverse direction, i.e. to derive the dynamics of the futures price process based on the dynamics of the risk process and the logspread. One can then proceed in a similiar manner.

The model introduced has some similarities with Gaussian commodity spot models, e.g. with the ones discussed in Schwartz [1997] or with the more general model provided in Casassus and Collin-Dufresne [2005]. In these models the triple of futures log price, spot log price and logspread is a 3-dimensional Gaussian process, too. These spot models, however, have different aims; e.g. they can be used for pricing long term forward commitments on the *same* commodity. Any forward position can be hedged by using one interest rate derivative and two short term futures contracts. This means that the latter models are *complete* and hence the model dynamics under the *risk-neutral measure* have to be calibrated to current futures and derivative prices.

The main aim of our model instead is to analyze the hedge error entailed when cross hedging risk exposures with futures written on a correlated, but *different* risk source. Our model includes a non-hedgeable risk factor, the spread, leading to *incompleteness*. We work under the physical measure since this is the only measure under which the hedge errors characteristics are relevant for risk management. Moreover, in a cross hedging situation a calibration is not always possible, e.g. if there are no liquid kerosene futures.

2.1.2. An Empirical Illustration

In the following we illustrate the estimation of the model by reconsidering the example of kerosene and crude oil. We use daily data of the spot kerosene price and the price of

Table 2.1.: Estimates

contract	nobs	ADF	μ	σ_x	σ_s	κ	m	ρ
201010	1147	0.0888*	0.0751	0.2768	0.2871	2.5548	−0.1661	0.3401
			(0.1302)	(0.0059)	(0.0060)	(1.0976)	(0.0534)	(0.0260)
201009	1147	0.0750*	0.0742	0.2795	0.2877	2.7131	−0.1687	0.3475
			(0.1324)	(0.0059)	(0.0060)	(1.1078)	(0.0515)	(0.0260)
201008	1145	0.0631*	0.0763	0.2823	0.2886	2.8890	−0.1693	0.3550
			(0.1361)	(0.0059)	(0.0061)	(1.1728)	(0.0492)	(0.0258)
201007	1123	0.0555*	0.0778	0.2855	0.2908	3.0632	−0.1731	0.3646
			(0.1348)	(0.0061)	(0.0062)	(1.1704)	(0.0473)	(0.0259)
201006	1375	0.0310**	0.1703	0.2758	0.2993	2.5899	−0.1946	0.3404
			(0.1209)	(0.0053)	(0.0058)	(1.1150)	(0.0493)	(0.0241)
201005	1080	0.0436**	0.1114	0.2901	0.2942	3.5046	−0.1752	0.3771
			(0.1405)	(0.0064)	(0.0064)	(1.2913)	(0.0410)	(0.0262)
201004	1058	0.0326**	0.0906	0.2954	0.2970	3.8040	−0.1808	0.3867
			(0.1453)	(0.0065)	(0.0065)	(1.3324)	(0.0390)	(0.0262)
201003	1035	0.0250**	0.0737	0.3002	0.3006	4.1174	−0.1836	0.3990
			(0.1494)	(0.0066)	(0.0067)	(1.4241)	(0.0371)	(0.0262)
201002	1015	0.0193**	0.0906	0.3035	0.3020	4.5175	−0.1866	0.4043
			(0.1518)	(0.0068)	(0.0069)	(1.6076)	(0.0332)	(0.0263)
201001	994	0.0126**	0.0789	0.3098	0.3045	4.9939	−0.1916	0.4178
			(0.1551)	(0.0070)	(0.0069)	(1.3783)	(0.0312)	(0.0262)
200912	1245	0.0097***	0.1852	0.2979	0.3114	3.8232	−0.2137	0.3892
			(0.1345)	(0.0060)	(0.0064)	(1.2755)	(0.0374)	(0.0241)
200911	950	0.0046***	0.0860	0.3197	0.3108	6.3841	−0.1994	0.4415
			(0.1564)	(0.0073)	(0.0072)	(2.0112)	(0.0238)	(0.0261)
200910	928	0.0029***	0.0603	0.3229	0.3135	7.0739	−0.2032	0.4532
			(0.1687)	(0.0075)	(0.0074)	(1.9253)	(0.0231)	(0.0261)
200909	906	0.0014***	0.0819	0.3275	0.3178	8.0814	−0.2061	0.4677
			(0.1738)	(0.0077)	(0.0076)	(2.1057)	(0.0208)	(0.0260)
200908	885	≤ 0.0010***	0.0430	0.3321	0.3223	9.5437	−0.2120	0.4806
			(0.1807)	(0.0079)	(0.0078)	(2.2822)	(0.0178)	(0.0259)
200907	862	≤ 0.0010***	0.0743	0.3388	0.3275	11.5957	−0.2166	0.5011
			(0.1828)	(0.0082)	(0.0080)	(2.5298)	(0.0153)	(0.0256)
200906	1114	≤ 0.0010***	0.1416	0.3191	0.3298	6.6833	−0.2404	0.4579
			(0.1498)	(0.0068)	(0.0072)	(1.7074)	(0.0233)	(0.0239)
200905	819	≤ 0.0010***	−0.0114	0.3516	0.3407	15.7614	−0.2274	0.5364
			(0.1953)	(0.0086)	(0.0086)	(2.9919)	(0.0111)	(0.0250)
200904	797	≤ 0.0010***	−0.0655	0.3513	0.3429	15.9689	−0.2307	0.5594
			(0.1973)	(0.0088)	(0.0087)	(3.0523)	(0.0103)	(0.0245)
200903	775	≤ 0.0010***	−0.0672	0.3436	0.3351	15.0723	−0.2362	0.5631
			(0.1953)	(0.0086)	(0.0086)	(2.8192)	(0.0135)	(0.0244)
200902	755	≤ 0.0010***	−0.0694	0.3469	0.3257	12.1794	−0.2411	0.5815
			(0.2005)	(0.0089)	(0.0085)	(2.4786)	(0.0146)	(0.0242)
200901	733	0.0021***	−0.0802	0.3306	0.3053	10.9528	−0.2398	0.5742
			(0.1966)	(0.0086)	(0.0081)	(2.4631)	(0.0160)	(0.0247)

The first column presents the maturity date of the contract, the second column gives the number of observations. The third column reports the p-value of the augmented Dickey-Fuller test for the null of non-stationarity. The * (**,***) indicates the rejection of non-stationary at the 10% (5%,1%) level. The remaining columns show the parameter estimates with the corresponding asymptotic standard errors given in parenthesis.

different crude oil futures contracts. The maturities of the futures contracts range from January 2009 until October 2010, resulting in 21 (overlapping) time series.

In a first step we check via the augmented Dickey-Fuller test whether the logspread of the kerosene spot price and the corresponding crude oil futures price is stationary. Table 2.1 reports the results for the different futures contracts. For most of the pairs we reject the null of a non-stationary logspread at any reasonable level. Of course, as one would expect for a statistical test, the procedure does not suggest the existence of a long-term relationship for every pair even if it is present. For these cases the model uncertainty is obvious and we will check in Section 2.4.1 how the application of an optimal hedging strategy influences the hedging performance if the strategy is derived under our model but is applied to the 2GBM model, and vice versa.

Table 2.1 also presents the estimation result for our data sets. To concentrate on one asset in the remaining part of the chapter we use one *representative contract* with moderate, not extreme, parameter values especially for κ and ρ. We choose the contract with maturity in August 2009. The number of observations at which both assets, the futures contract and the spot kerosene, are traded is 885. Figure 2.1 shows the time evolvement of these two price series. Obviously, the price evolvements are very similar, which is also supported by the time-series plot of the logspread of the log prices (depicted in the lower panel).

In the next sections we derive the variance optimal hedge, the corresponding variance of the hedge error and derive quantitative and qualitative statements in terms of the structural parameters.

2.2. Optimal Variance Hedging with Futures Contracts

Suppose that a hedger sets up a portfolio consisting of futures contracts and cash positions, in order to hedge the risk position $h(I_T)$. In the following we denote by ξ_t the number of futures contracts held in the portfolio at time t. We assume that any futures position strategy $\xi = (\xi_t)_{t\in[0,T]}$ is non-anticipating, i.e. at any time it incorporates only information publicly available. In mathematical terms, this means that ξ is progressively measurable with respect to $(\mathcal{F}_t)_{t\geq 0}$, the filtration generated by the Brownian motions $\left(W^{(X)},\ W^{\perp}\right)^{\mathsf{T}}$ and completed with the P-null sets of the basis.

If the futures price changes by ΔX_t from one trading day to the next, the hedger's margin account is adjusted by ΔX_t per futures contract. The cash position in the hedging portfolio is changed accordingly, entailing a portfolio value change due to variation margins of $\Delta V_t^{\text{mar}} = \xi_t \Delta X_t$.

Denote by $V = (V_t)_{t\in[0,T]}$ the *total value* of the hedging portfolio. Given an interest rate r, the cash position contributes to the portfolio by $rV_t dt$, hence the total value satisfies the continuous-time self-financing condition

$$dV_t = \xi_t dX_t + rV_t dt. \tag{2.7}$$

Equation (2.7) is linear. Given an initial portfolio value of $V_0 = v$, the portfolio process

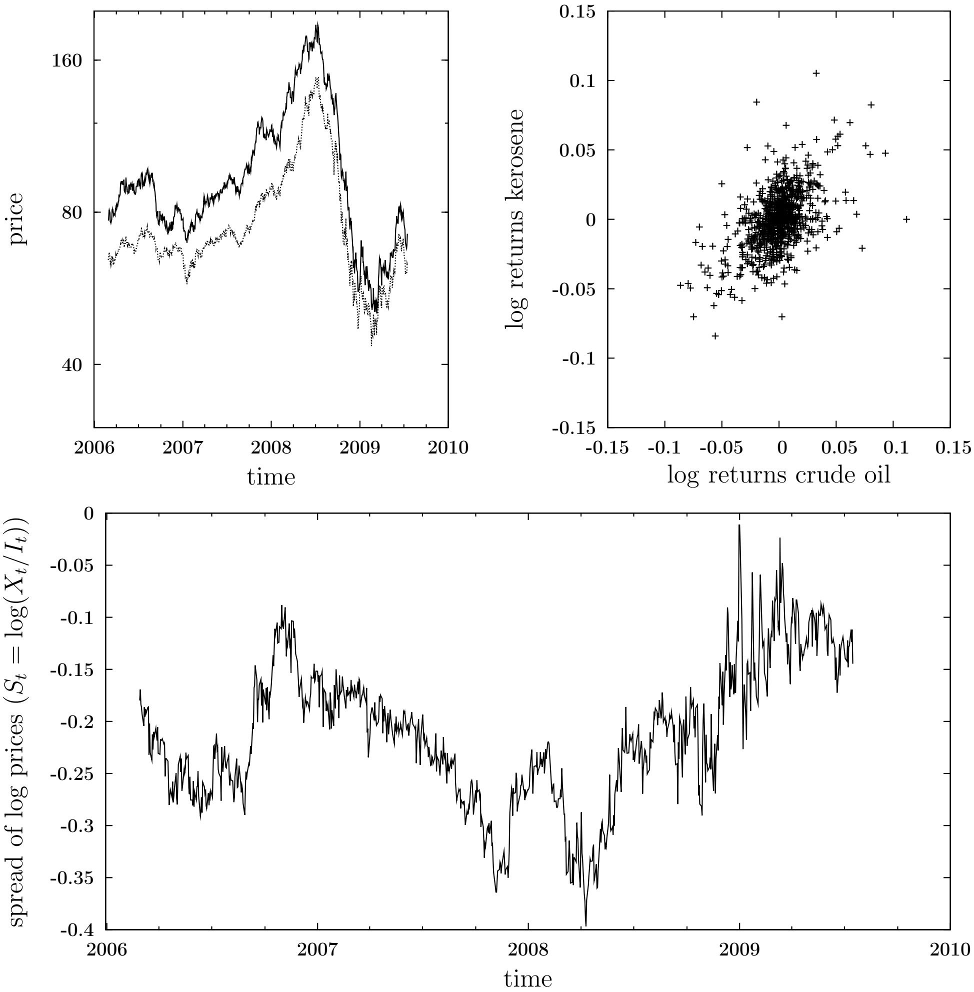

Figure 2.1.: The upper left hand panel depicts the time evolvement of the daily price of the crude oil futures with maturity in August 2009 in US$/BBL (dashed line) and for spot jet kerosene in US$/BBL (solid line) from 2006/02/27 until 2009/07/16 (resulting in 885 observations). The structure of this figure is the same as Figure 0.2 on page 10. The upper right hand panel exhibits the scatter plot of the daily returns. The lower panel depicts the time evolvement of the logspread of the log prices.

has the explicit representation

$$V_t = e^{rt}\left(v + \int_0^t e^{-rs}\xi_s dX_s\right).$$

Consider a self-financing hedge portfolio with futures position ξ_t at time t. The *conditional hedge error* of the portfolio at time $t \in [0,T]$ is then given by

$$\begin{aligned} C_t(\xi, v) &= E\left[e^{-r(T-t)}h(I_T)|\mathcal{F}_t\right] - V_t \qquad (2.8)\\ &= e^{rt}\left(E\left(e^{-rT}h(I_T)|\mathcal{F}_t\right) - v - \int_0^t \xi_s e^{-rs} dX_s\right). \end{aligned}$$

$C_T(\xi, v)$ will also be referred to as the *realized hedge error*. Note that if $C_t(\xi, v)$ is negative, the combined value of the risk and the hedge portfolio is expected to end up with a plus.

To determine the variance optimal strategy within our model, i.e. the strategy minimizing the variance of the realized hedge error, we suppose that X is a martingale. This is a plausible assumption since the empirical analysis of crude oil futures prices shows that the estimated drift parameter is close to zero for all contract months and statistically insignificant for most assets (see Table 2.1 in Section 2.1.2). In addition, as the estimation of the drift is notoriously challenging and very often highly speculative, it can easily distort the main aim of hedging, which is the reduction of risk. We therefore discuss here the martingale case in depth and postpone the discussion of the more general case to Section 2.5.

Assuming that X is a martingale means that $\mu = 0$ and $dX_t = \sigma_X X_t dW_t^{(X)}$. Then Equation (2.8) implies that the discounted conditional hedge error is also a martingale. The martingale $e^{-rT}E(h(I_T)|\mathcal{F}_t)$ can be written as

$$e^{-rT}E\left(h(I_T)|\mathcal{F}_t\right) = e^{-rT}E(h(I_T)) + \int_0^t a_s dW_s^{(X)} + \int_0^t b_s dW_s^{\perp}, \quad t \in [0,T], \qquad (2.9)$$

where a and b are progressively measurable and square-integrable processes. The first stochastic integral on the right hand side is hedgeable, since it is driven by the same BM as the futures X. More precisely, following the strategy

$$\xi_t^* = \frac{a_t e^{rt}}{\sigma_X X_t}, \qquad (2.10)$$

the gain from the futures position up to time t satisfies $\int_0^t \xi_s^* e^{-rs} dX_s = \int_0^t a_s dW_s^{(X)}$. The second integral in Equation (2.9) is orthogonal to $W^{(X)}$, and hence completely non-hedgeable with X. This implies that the strategy ξ^* minimizes the variance of the realized hedge error (see Theorem 1 in Föllmer and Sondermann [1986] where this argument has been employed for the first time). Moreover, the conditional hedge error satisfies

$$C_t(\xi^*, v) = e^{rt} \int_0^t b_s dW_s^{\perp} + E\left(e^{-(T-t)r} h(I_T)\right) - e^{rt} v.$$

Profiting from the Markov property of the processes I and S, we may express a and b in terms of sensitivities of the expected risk with respect to the futures and the logspread values. More precisely, let

$$\psi(t, x, s) = e^{-r(T-t)} E\left(h\left(X_T^{t,x} e^{-S_T^{t,s}}\right)\right),$$

where $X^{t,x}$ and $S^{t,s}$ are the solutions of Equation (2.1) resp. Equation (2.2) on $[t, T]$ with initial conditions $X_t^{t,x} = x$ resp. $S_t^{t,s} = s$. We will refer to ψ as the *value function*. If h is Lipschitz continuous and its weak derivative h' is Lebesgue-almost everywhere differentiable, then ψ is continuously differentiable with respect to x and s, and

$$\psi_x(t, x, s) = \frac{\partial}{\partial x}\psi(t, x, s) = e^{-r(T-t)} E\left(h'\left(X_T^{t,x} e^{-S_T^{t,s}}\right) X_T^{t,1} e^{-S_T^{t,s}}\right). \tag{2.11}$$

For details we refer the interested reader to Section 3.3.3, where a similar a statement is shown. Notice that $\partial S_T^{t,s}/\partial s = e^{-\kappa(T-t)}$, and hence by the same reasoning

$$\begin{aligned}\psi_s(t, x, s) = \frac{\partial}{\partial s}\psi(t, x, s) &= -e^{-r(T-t)} E\left(h'\left(X_T^{t,x} e^{-S_T^{t,s}}\right) X_T^{t,x} e^{-S_T^{t,s}} e^{-\kappa(T-t)}\right) \\ &= -e^{-\kappa(T-t)} x \frac{\partial}{\partial x}\psi(t, x, s).\end{aligned} \tag{2.12}$$

The pair $(X, S)^{\mathsf{T}}$ is a 2-dimensional SDE, driven by $\left(W^{(X)}, W^{\perp}\right)^{\mathsf{T}}$ via the diffusion matrix

$$\tilde{\Sigma}(x, s) = \begin{pmatrix} \sigma_X x & 0 \\ \rho\sigma_S & \bar{\rho}\sigma_S \end{pmatrix}.$$

With Itō's formula we obtain that the processes a and b, appearing in the martingale representation (2.9), are given by

$$\begin{pmatrix} a_t \\ b_t \end{pmatrix} = e^{-rt}\, \tilde{\Sigma}^{\mathsf{T}}(X_t, S_t) \begin{pmatrix} \psi_x(t, X_t, S_t) \\ \psi_s(t, X_t, S_t) \end{pmatrix}.$$

From Equation (2.10) and Equation (2.12) we can deduce the following result describing the variance optimal hedge in terms of the futures-Delta, and the minimal hedge error in terms of the logspread-Delta.

Theorem 2.2. *The variance optimal futures position in the hedging portfolio is given by*

$$\xi_t^* = \left[1 - \frac{\sigma_S}{\sigma_X}\rho e^{-\kappa(T-t)}\right] \psi_x(t, X_t, S_t), \tag{2.13}$$

and entails a realized hedge error of

$$C_T(\xi^*, v) = E(h(I_T)) - e^{rT}\left(v - \bar{\rho}\int_0^T e^{-rt}\sigma_S\psi_s(t, X_t, S_t)dW_t^{\perp}\right).$$

Observe that the optimal hedge ξ^* is the Delta of the position's expectation, dampened by the *hedge ratio* defined by

$$f(T-t) = 1 - \frac{\sigma_S}{\sigma_X}\rho e^{-\kappa(T-t)}.$$

The factor essentially equals 1 if the product of time to maturity $T-t$ and reversion speed κ is large. In this case the logspread is expected to return to its mean reversion level before maturity, and hence the optimal hedge ratio equals one. As maturity approaches, the short-term fluctuations have an increasing impact on the terminal hedge performance, making the hedge ratio converge to

$$h = \rho_{IX}\frac{\sigma_I}{\sigma_X}. \tag{2.14}$$

Indeed, due to Equation (2.3), we have $\lim_{(T-t)\downarrow 0} 1 - \sigma_S\rho e^{-\kappa(T-t)}/\sigma_X = 1 - \sigma_S\rho/\sigma_X = h$. We remark that h, defined in Equation (2.14), is sometimes referred to as the *minimum variance hedge ratio* (see e.g. Chapter 3.4 in Hull [2008]).

Observe that if κ is equal to zero, which essentially means that there is no mean reversion, then the logspread is not stationary and its variance increases linearly with time. The hedge ratio is not dampened and it coincides with h. In this case the strategy ξ^* is equal to the optimal strategy in a model where both X and I are modeled as GBMs (see Section 2.4.1 for more details).

In Formula (2.13) the optimal hedge is expressed in terms of the Delta with respect to the futures price. In order to obtain a representation in terms of the Delta with respect to the illiquid asset price I, define first $\varphi(t, y, s) = e^{-r(T-t)}E(h(I_T^{t,y,s}))$, where $I^{t,y,s}$ is the solution of the SDE for the illiquid asset on $[t, T]$, with initial values $I_t^{t,y,s} = y$ and $S_t^{t,s} = s$. Note that $\psi(t, x, s) = \varphi(t, e^{-s}x, s)$, and in particular, $\psi_x(t, x, s) = e^{-s}\varphi_y(t, e^{-s}x, s)$. Thus, the optimal hedge may be rewritten as

$$\xi_t^* = e^{-S_t}\left[1 - \frac{\sigma_S}{\sigma_X}\rho e^{-\kappa(T-t)}\right]\varphi_y(t, I_t, S_t). \tag{2.15}$$

If the logspread is positive, then the illiquid asset price is expected to rise relative to the futures price. This explains why in Equation (2.15) the Delta is reduced by the factor e^{-S_t}. Conversely, if the logspread is negative, then the illiquid asset is expected to fall relative to the futures price. In this the case the Delta is augmented by the factor e^{-S_t}.

Finally, we remark that the hedge ratio remains the same if the hedger uses an option on the futures for hedging the risk exposure $h(I_T)$. Denote by $\Delta(t, x)$ the option's Delta at time t, given a futures price of $X_t = x$. The dynamics of the option price $P(t, X_t)$ satisfy $dP(t, X_t) = rP(t, X_t)dt + \Delta(t, X_t)dX_t$, and the value of a self-financing portfolio containing ξ_t options at time t is given by

$$V_t = e^{rt}\left(V_0 + \int\limits_0^t e^{-rs}\xi_s\Delta(s, X_s)dX_s\right).$$

The variance minimizing *option position* can be shown to be equal to

$$\xi_t^* = \left[1 - \frac{\sigma_S}{\sigma_X}\rho e^{-\kappa(T-t)}\right]\frac{\psi_x(t, X_t, S_t)}{\Delta(t, X_t)}.$$

Suppose that a non-linear position of kerosene is hedged with an option, written on a crude oil futures, having a similar payoff profile. Then the ratio $\psi_x(t, X_t, S_t)/\Delta(t, X_t)$ is usually stable and hence the hedging portfolio does not need to be rebalanced as strongly as when using futures for hedging.

2.3. Standard Deviation of the Hedge Error

Having derived the variance optimal strategy and the corresponding hedge error, see Theorem 2.2, we now aim at computing the implied standard deviation of the hedge error. This allows us to quantify the risk associated with the optimal strategy, which is important for risk management and performance evaluation of the hedging strategy. We therefore derive analytic and semianalytic formulas for the standard deviation of the hedge error when minimizing the variance of risk exposures within our model. As in the previous section, we assume that the hedger does not have any directional view concerning the futures. This means that the futures price X is a martingale with dynamics $dX_t = \sigma_X X_t dW_t$.

We aim at computing the standard deviation of the hedge error when cross-hedging the position $h(I_T)$ following the strategy ξ^* of Equation (2.13). Note that the standard deviation of $C_T(\xi^*, v)$ coincides with the standard deviation of
$e^{rT}\bar{\rho}\int_0^T e^{-rt}\sigma_S\psi_s(t, X_t, S_t)dW_t^\perp$. The Itô isometry implies

$$\operatorname{std}(C_T(\xi^*, v)) = e^{rT}\bar{\rho}\sigma_S\sqrt{\int\limits_0^T e^{-2rt}E[\psi_s^2(t, X_t, S_t)]dt}. \tag{2.16}$$

In general, there is no closed-form expression for the formula for the integral in Equation (2.16). For linear positions, however, we may explicitly calculate the variance, since the futures and the spread are lognormally distributed. Besides, for positions corresponding to Plain Vanilla options, the logspread-Delta ψ_s has an explicit representation, and thus allows for an efficient Monte Carlo simulation of the error (2.16). We proceed by discussing both cases separately.

2.3.1. Linear Positions

In this subsection we derive an analytic formula for the hedge error variance when cross-hedging a linear position. This is the most relevant case, since most of the risk positions of industrial companies are linear. Think for instance of an airline being exposed to a short position of kerosene.

Suppose that the payoff function h is given by $h(y) = cy$, with $c \in \mathbb{R}$. In this case the Delta of the value function with respect to the futures price satisfies $\psi_x(t, x, s) = e^{-r(T-t)} E(cX_T^{t,1} e^{-S_T^{t,s}})$ (see Equation (2.11)). Thus, with Equation (2.12), we get

$$\frac{\partial}{\partial s}\psi(t, x, s) = -e^{-\kappa(T-t)} x e^{-r(T-t)} E(cX_T^{t,1} e^{-S_T^{t,s}}).$$

In the following we do not only need to compute the expectation of the product $X_T^{t,1} e^{-S_T^{t,s}}$, but also the expectation of the product of higher moments of the logspread and the illiquid asset. Therefore we provide the following lemma.

Lemma 2.3. *Let $a \in \mathbb{R}$ and $b \in \mathbb{R}_+$, then*

$$E\left(e^{-aS_t^{0,s}}\left(X_t^{0,x}\right)^b\right) = A(a, b, x, s, t),$$

where $A(a, b, x, s, t)$ is defined by

$$\begin{aligned} A(a, b, x, s, t) \quad &= \\ x^b \exp&\left[-\frac{1}{2}\sigma_X^2 t\left(b - b^2\right) - ase^{-\kappa t} - a\left(m + b\rho\sigma_X\sigma_S\frac{1}{\kappa}\right)\left(1 - e^{-\kappa t}\right)\right] \\ &\times \exp\left[\frac{1}{2}a^2\sigma_S^2\frac{1}{2\kappa}\left(1 - e^{-2\kappa t}\right)\right]. \end{aligned} \tag{2.17}$$

Proof. Since

$$S_t^{0,s} = se^{-\kappa t} + m(1 - e^{-\kappa t}) + \int_0^t e^{-\kappa(t-u)}\sigma_S(\rho dW_u^{(X)} + \bar{\rho} dW_u^{\perp}),$$

we get

$$\begin{aligned} &e^{-aS_t^{0,s}}(X_t^{0,x})^b = \\ &x^b \exp\left(-\frac{b}{2}\sigma_X^2 t - ase^{-\kappa t} - am(1 - e^{-\kappa t}) + \int_0^t (b\sigma_X - \rho a\sigma_S e^{-\kappa(t-u)})dW_u^{(X)}\right) \\ &\qquad \times \exp\left(-\int_0^t (a\bar{\rho}\sigma_S e^{-\kappa(t-u)})dW_u^{\perp}\right). \end{aligned}$$

We calculate the variances of the independent normal variables given by the integrals in the last two factors. We have

$$\int_0^t (b\sigma_X - \rho a\sigma_S e^{-\kappa(t-u)})^2 dt = b^2\sigma_X^2 t - 2ab\rho\sigma_X\sigma_S\frac{1}{\kappa}(1-e^{-\kappa t}) + a^2\rho^2\sigma_S^2\frac{1}{2\kappa}(1-e^{-2\kappa t})$$

and

$$\int_0^t (a\bar{\rho}\sigma_S e^{-\kappa(t-u)})^2 dt = a^2\bar{\rho}^2\sigma_S^2\frac{1}{2\kappa}(1-e^{-2\kappa t}).$$

We use this to derive

$$\begin{aligned}
E\left(e^{-aS_t^{0,s}}(X_t^{0,x})^b\right) = {} & x^b \exp\left(-\frac{b}{2}\sigma_X^2 t - ase^{-\kappa t} - am(1-e^{-\kappa t})\right) \\
& \times E\left(\exp\left(\int_0^t (b\sigma_X - \rho a\sigma_S e^{-\kappa(t-u)}) dW_u^{(X)}\right)\right) \\
& \times E\left(\exp\left(-\int_0^t (a\bar{\rho}\sigma_S e^{-\kappa(t-u)}) dW_u^{\perp}\right)\right) \\
= {} & x^b \exp\left(-\frac{b}{2}\sigma_X^2 t - ase^{-\kappa t} - am(1-e^{-\kappa t})\right) \\
& \times \exp\left[\frac{1}{2}\left(b^2\sigma_X^2 t - 2ab\rho\sigma_X\sigma_S\frac{1}{\kappa}(1-e^{-\kappa t})\right)\right] \\
& \times \exp\left[\frac{1}{2}\left(a^2\rho^2\sigma_S^2\frac{1}{2\kappa}(1-e^{-2\kappa t}) + a^2\bar{\rho}^2\sigma_S^2\frac{1}{2\kappa}(1-e^{-2\kappa t})\right)\right],
\end{aligned}$$

from which the result follows. □

With this at hand the standard deviation in (2.16) simplifies to

$$\operatorname{std}(C_T(\xi^*, v)) = |c|\bar{\rho}\sigma_S\sqrt{\int_0^T e^{-2\kappa(T-t)} E\left[X_t^2 A^2(1,1,1,S_t,T-t)\right] dt}. \tag{2.18}$$

From this we are able to derive the following explicit formula for the hedge error variance.

Theorem 2.4. *The variance optimal cross-hedge of a linear position cI_T entails a hedge* *error with standard deviation*

$$
\begin{aligned}
&\operatorname{std}(C_T) = \\
&\sigma_S\sqrt{1-\rho^2}x\exp\left((m-s)e^{-\kappa T} - m - \rho\sigma_X\sigma_S\frac{1}{\kappa}\left(1-2e^{-\kappa T}\right) + \sigma_S^2\frac{1}{4\kappa}\left(1-2e^{-\kappa T}\right)\right) \\
&\qquad\times |c|\sqrt{\int_0^T \exp\left(-2\kappa(T-t) + \sigma_X^2 t - 2\rho\sigma_X\sigma_S\frac{1}{\kappa}e^{-\kappa(T-t)} + \sigma_S^2\frac{1}{2\kappa}e^{-2\kappa(T-t)}\right)dt}.
\end{aligned}
\tag{2.19}
$$

Proof. Recall that from Equation (2.18) we have

$$
\operatorname{Var}\left(C_T(\xi^*, v)\right) = c^2\bar{\rho}^2\sigma_S^2\int_0^T e^{-2k(T-t)}E\left(X_t^2A^2(1,1,1,S_t,T-t)\right)dt. \tag{2.20}
$$

By the definition of A, we have

$$
\begin{aligned}
&E\left(X_t^2A^2(1,1,1,S_t,T-t)\right) = \\
&\qquad\exp\left(-2(m+\rho\sigma_X\sigma_S\frac{1}{\kappa})(1-e^{-\kappa(T-t)}) + \sigma_S^2\frac{1}{2\kappa}(1-e^{-2\kappa(T-t)})\right) \\
&\qquad\qquad\times E\left(X_t^2\exp\left(-2S_te^{-\kappa(T-t)}\right)\right),
\end{aligned}
\tag{2.21}
$$

and again using the definition of A we get

$$
\begin{aligned}
E\left(X_t^2\exp\left(-2S_te^{-\kappa(T-t)}\right)\right) &= A(2e^{-\kappa(T-t)}, 2, x, s, t) \\
&= x^2\exp\left[-2(m+2\rho\sigma_X\sigma_S\frac{1}{\kappa})(e^{-\kappa(T-t)} - e^{-\kappa T})\right] \\
&\quad\times\exp\left[\sigma_X^2t - 2se^{-\kappa T} + 2\sigma_S^2\frac{1}{2\kappa}(e^{-2\kappa(T-t)} - e^{-2\kappa T})\right].
\end{aligned}
$$

Combining the last equation with Equation (2.21) we further obtain

$$
\begin{aligned}
&E\left(X_t^2A^2(1,1,1,S_t,T-t)\right) = \\
&\qquad x^2\exp\left(\sigma_X^2t - 2se^{-\kappa T} - 2\rho\sigma_X\sigma_S\frac{1}{\kappa}(1+e^{-\kappa(T-t)} - 2e^{-\kappa T})\right) \\
&\qquad\qquad\times\exp\left(-2m(1-e^{-\kappa T}) + \sigma_S^2\frac{1}{2\kappa}(1-2e^{-\kappa T} + e^{-2\kappa(T-t)})\right).
\end{aligned}
\tag{2.22}
$$

The previous calculations yield, by combination of Equation (2.20) and Equation (2.22), Expression (2.19). □

The integral in Expression (2.19) can be computed in a straightforward manner using standard numerical quadratures algorithms.

When analyzing the dependence of the hedge error on the different model parameters it is convenient to rewrite the hedge error formula (2.19) as follows:

$$\begin{aligned}\operatorname{std}(C_T) &= |c|\,\sigma_S\sqrt{1-\rho^2}x\exp\left((m-s)e^{-\kappa T}-m\right)\left[\int_0^T \exp\left(-2\kappa(T-t)+\sigma_X^2 t\right)\right.\\ &\left.\times\exp\left(-2\rho\sigma_X\frac{\sigma_S}{\kappa}\left(1+e^{-\kappa(T-t)}-2e^{-\kappa T}\right)+\frac{\sigma_S^2}{2\kappa}\left(1+e^{-2\kappa(T-t)}-2e^{-\kappa T}\right)\right)dt\right]^{\frac{1}{2}}.\end{aligned}$$

The hedge error (2.19) can be approximated with simplified formulas. For long maturities, i.e. large T, the factor $\exp\big((m-s)e^{-\kappa T}-m\big)$ approximately coincides with e^{-m}. Moreover,

$$\int_0^T e^{-2\kappa(T-t)+\sigma_X^2 t}dt = \frac{e^{\sigma_X^2 T}-e^{-2\kappa T}}{2\kappa+\sigma_X^2} \approx \frac{e^{\sigma_X^2 T}}{2\kappa+\sigma_X^2},$$

and hence

for long maturities: $$\operatorname{std}(C_T) \approx |c|\,xe^{-m}\sigma_S\sqrt{1-\rho^2}e^{\left(-\rho\sigma_S\sigma_X+\frac{\sigma_S^2}{4}\right)/\kappa}\sqrt{\frac{e^{\sigma_X^2 T}}{2\kappa+\sigma_X^2}}.$$

Observe that the hedge error increases exponentially with time to maturity. However, the volatility squared σ_X^2 is usually low (see Table 2.1 in Section 2.1.2), and thus the hedge error increases approximately linearly, with slope $\sigma_X^2/2$, in the first several years (compare with the upper left panel in Figure 2.2).

For short maturities, i.e. small T, the factor $\exp\big((m-s)e^{-\kappa T}-m\big)$ approximately coincides with e^{-s}. Besides, by linearly approximating exponentials with the Taylor expansion up to the first order, we have

$$\int_0^T e^{-2\kappa(T-t)+\sigma_X^2 t}dt = \frac{e^{-\sigma_X^2 T}-e^{-2\kappa T}}{2\kappa+\sigma_X^2} \approx T.$$

Therefore, we get the approximation

for short maturities: $$\operatorname{std}(C_T) \approx |c|\,\sigma_S\sqrt{1-\rho^2}xe^{-s}\sqrt{T}.$$

Note that the hedge error increases with order $T^{1/2}$, for short maturities (see the upper left panel in Figure 2.2). The kink in Figure 2.2 is determined by how fast the Ornstein Uhlenbeck process describing the logspread attains its stationary distribution. The variance of the logspread at time T, as a function of time to maturity $T-t$, is given by $\sigma_S^2(1-e^{-2\kappa(T-t)})/(2\kappa)$. The hedge error is roughly proportional to the variance of the logspread.

The hedge error vanishes as the variance of the stationary logspread distribution,

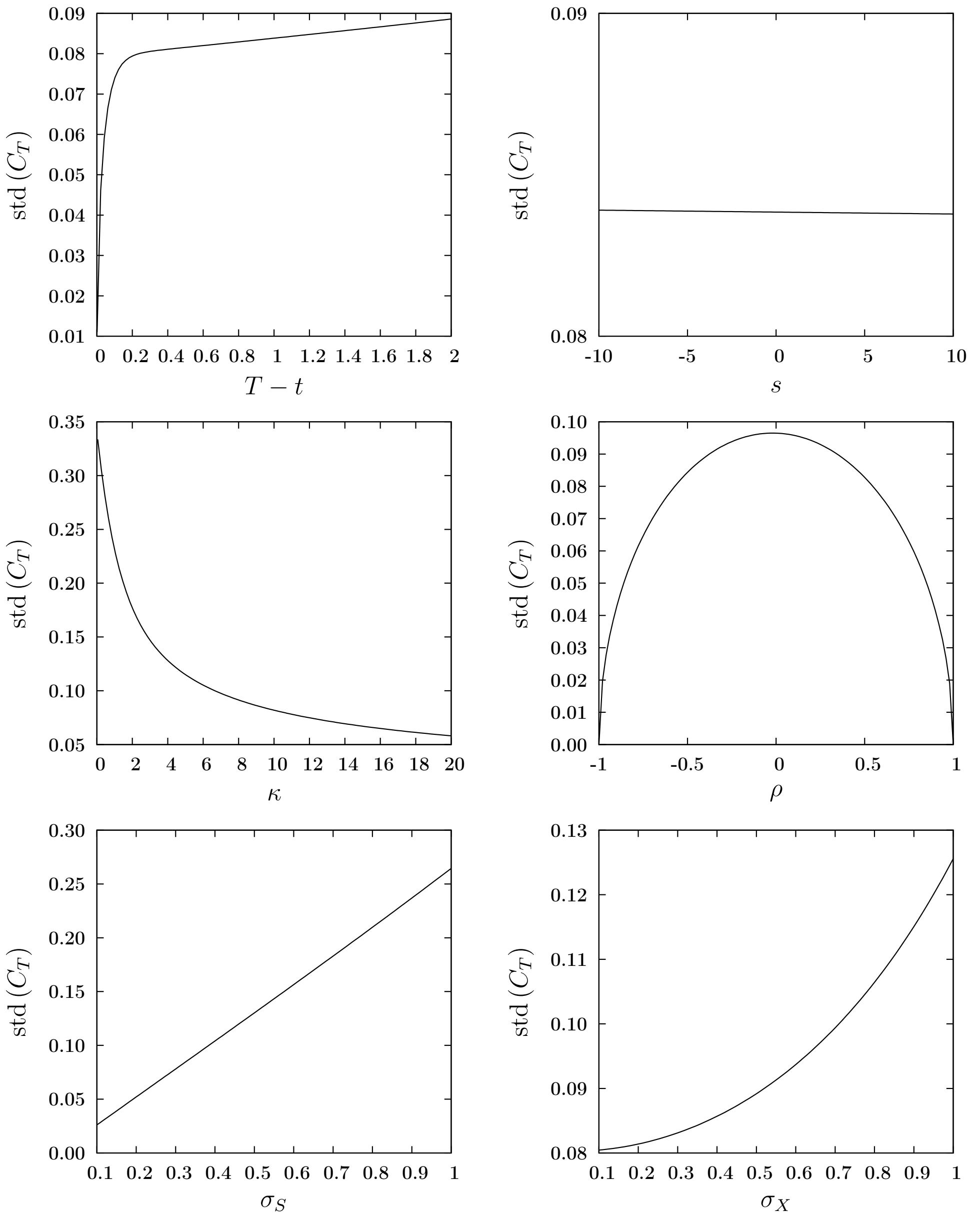

Figure 2.2.: This figure shows the sensitivity of the standard deviation of the hedge error with respect to the time to maturity (in the upper left hand panel) and with respect to the parameters of the model (remaining panels). In each panel only the parameter indicated on the abscissa is varied while the others remain fixed at the estimates from the futures contract with maturity in August 2009.

$\sigma_S^2/(2\kappa)$, converges to zero. Moreover, it is straigthtforward to show

$$\lim_{\kappa\downarrow 0} \operatorname{std}(C_T) = |c|\, xe^{-s} \frac{\sigma_S}{\sigma_X} \sqrt{1-\rho^2} \sqrt{e^{\sigma_X^2 T} - 1}.$$

Of course, not only the impact of the time to maturity to the hedge error is of interest but also the influence of the model's parameters. Figure 2.2 highlights the sensitivity of the standard deviation of the hedge error towards changes in the parameters. The figure depicts the resulting standard deviation by changing one parameter and keeping the others constant. The fixed parameters are set to the estimates of the futures contract with maturity in August 2009. The figure shows (in the middle left panel) the decreasing standard deviation in the mean reversion speed κ. This comes at no surprise as a larger κ results in a faster return to the long-term relationship. The reverse U-shaped behavior with respect to the correlation ρ (in the right middle panel) highlights the change from the incomplete market setting for $|\rho| < 1$ to the complete market setting for $|\rho| = 1$. With increasing instantaneous variance of the logspread and the futures price process (σ_S and σ_X) the variance of the hedge error also increases (shown in the lower panels).

2.3.2. Non-linear Positions

In practice the case of non-linear risk positions is also relevant. For instance, consider the very illiquid German natural gas futures markets. Due to the illiquidity gas traders frequently use futures of neighbouring countries for hedging purposes. So, if operators of German gas power plants protect themselves against changing gas prices by buying Dutch gas on the futures market (e.g. natural gas futures of the Dutch market TTF) the basis risk is due to a geographical spread in commodity prices which arises from different trading places for the same underlying. In this case TTF contracts serve as proxies that are cointegrated with the natural gas prices in the German market area. The profit margin of a gas power plant is essentially determined by the spark spread, i.e. the spread between the electricity price per MWh and the price of the amount of gas the plant needs for producing 1 MWh of electricity. Electricity will only be produced if the profit margin exceeds the operating costs. A gas power plant can thus be seen as a call option, a highly non-linear position, on the spark spread. We therefore also consider here the hedging of non-linear risk positions.

When it comes to hedging non-linear risk positions, the problem is that in general there are no explicit formulas available for the standard deviation of the hedge error. However, for Plain Vanilla options there are analytic formulas for the Deltas, allowing for swift simulation analysis. Indeed, the Deltas resemble the Deltas for Plain Vanilla options in the Black-Scholes model. As an example we provide the relevant formulas for a European call option with strike K, i.e. h is given by $h(y) = (y-K)^+$. First observe that the value function ψ for call options is given by

$$\psi(t,x,s) = E\left(\left(I_T^{t,x,s} - K\right)^+\right).$$

From the proof of Lemma 2.3 it can be seen that $I_T^{t,x,s} = e^{-S_T^{t,s}} X_T^{t,x}$ can be written as

$$I_T^{t,x,s} = x \exp\left(-\frac{1}{2}\sigma_X^2(T-t) - se^{-\kappa(T-t)} - m(1-e^{-\kappa(T-t)})\right)\exp(\sigma N),$$

where N is a standard normal variable and σ^2 is given by

$$\sigma^2 = \sigma_X^2(T-t) - 2\rho\sigma_X\sigma_S\frac{1}{\kappa}\left(1-e^{-\kappa(T-t)}\right) + \sigma_S^2\frac{1}{2\kappa}\left(1-e^{-2\kappa(T-t)}\right).$$

We get

$$\psi(t,x,s) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(xe^{-\frac{1}{2}\sigma_X^2(T-t)-se^{-\kappa(T-t)}-m(1-e^{-\kappa(T-t)})+\sigma y} - K\right)^+ e^{-y/2}dy. \tag{2.23}$$

In analogy to the standard Black-Scholes case we define the functions $d_+(t,x,s)$ and $d_-(t,x,s)$ as

$$d_+(t,x,s) = \frac{1}{\sigma}\left[\log\left(\frac{K}{x}\right) + \frac{1}{2}\sigma_X^2(T-t) + se^{-\kappa(T-t)} + m\left(1-e^{-\kappa(T-t)}\right)\right]$$

and $d_-(t,x,s) = d_+(t,x,s) - \sigma$.

Note that the integrand in the integral in Equation (2.23) equals zero if $y < d_+(t,x,s)$, and hence

$$\begin{aligned}\psi(t,x,s) = x\exp&\left(-\frac{1}{2}\sigma_X^2(T-t) - se^{-\kappa(T-t)} - m(1-e^{-\kappa(T-t)}) + \frac{1}{2}\sigma^2\right)\Phi(d_-(t,x,s))\\ &- K\Phi(d_+(t,x,s)),\end{aligned}$$

where Φ denotes the cumulative distribution function of the standard normal distribution. The above explicit representation of $\psi(t,x,s)$ yields

$$\begin{aligned}\psi_x(t,x,s) = \exp&\left(-\frac{1}{2}\sigma_X^2(T-t) - se^{-\kappa(T-t)} - m(1-e^{-\kappa(T-t)}) + \frac{1}{2}\sigma^2\right)\\ &\times\left[\Phi(d_-(t,x,s)) + \varphi(d_-(t,x,s))\partial_x d_-(t,x,s)\right]\\ &- K\varphi(d_+(t,x,s))\partial_x d_+(t,x,s),\end{aligned}$$

where $\partial_x d_+(t,x,s) = \partial_x d_-(t,x,s) = 1/(\sigma x)$.

The analytic and semianalytic formulas for the hedge error allow the efficient computation and the comparison of the hedge error variance for different futures contracts. Up to now, however, we assume that we hedge within the correct model (with an asymptotic stationary logspread). Although, statistical tests may help to decide whether the logspread is asymptotic stationary or not, there is always the risk to *hedge within the wrong model* and a relevant question arises: how sensitive is the hedge error with respect to the model choice? We address this question in the next Section.

2.4. The Performance of Suboptimal Hedging Strategies

So far we have assumed that we know with certainty that the price of the illiquid asset and the price of the liquid futures contract are cointegrated and evolve according to our model. However, it may happen that a statistical test leads to a wrong conclusion or different tests lead to different implications. In other words, we face model uncertainty.

In the following Section 2.4.1 we consider a 2GBM model and derive the hedge error obtained by using the optimal strategy from our model. Furthermore, we analyse the impact of applying the optimal hedging strategy from the 2GBM model to our model. We then proceed by comparing the optimal *dynamic* hedge with its optimal *static* counterpart. In practice, traders often hedge linear positions statically, holding a position in futures that corresponds to the size of the risk. By this they intuitively reflect that the hedge ratio essentially equals 1 whenever time to maturity is long. In Section 2.4.2 we first derive the optimal *static* hedging strategy and compare it with the hedging strategy ξ^*, which allows for portfolio regrouping.

2.4.1. The Costs of Ignoring a Long-term Relationship or Falsely Assuming a Long-term Relationship

We next introduce a simple model where both X *and* I are GBMs, hence are not cointegrated and the logspread does not have an asymptotic stationary distribution. We will refer to this model as the 2GBM model.

In both models, the futures price is assumed to satisfy the dynamics

$$dX_t = \sigma_X X_t dW_t^{(X)},$$

but in contrast to the model with a mean reverting logspread, discussed in the previous sections, the 2GBM model assumes that the illiquid asset price process is also a GBM with dynamics

$$dI_t = \sigma_I I_t \left(\rho_{IX} dW_t^{(X)} + \bar{\rho}_{IX} dW_t^{\perp} \right).$$

In this model the variance minimizing hedging strategy for European options with payoff function h are known to have the simple form

$$\zeta_t = \rho_{IX} \frac{\sigma_I I_t}{\sigma_X X_t} \psi_y(t, I_t), \tag{2.24}$$

where $\psi(t,y) = e^{-r(T-t)} E(h(I_T^{t,y}))$. For a derivation of Equation (2.24) we refer to Hulley and McWalter [2008]; see also Chapter 3 for a derivation in a slightly more general setting using BSDEs.

The optimal cross-hedge within the 2GBM model (2.24), is essentially determined by the cross correlation. If an airline company used a 2GBM model estimated with daily data to hedge kerosene short positions, then it would considerably underhedge its kerosene risk, facing thus unnecessarily high variations in costs. But, by how much does the hedge error increase if we use the wrong model? We next quantify the risk by calculating the hedge error when using the optimal strategy ζ of the 2GBM model while the log prices are cointegrated and evolve according to the dynamics of our model.

We restrict our analysis to linear positions of the form cI_T. As before we denote the realized hedge error by

$$C_T = cI_T - e^{rT}\left(v + \int_0^T e^{-rs}\zeta_s dX_s\right).$$

The following proposition provides the hedge error variance for the strategy (2.24) under our model with cointegration and under the 2GBM model.

Proposition 2.5. *Hedging the linear position cI_T with the strategy ζ entails a hedge error in the cointegration model with variance*

$$\begin{aligned}\operatorname{Var}(C_T) = c^2\Bigg\{& A(2,2,X_0,S_0,T) - A^2(1,1,X_0,S_0,T)\\ & -2\int_0^T \rho_{IX}\sigma_I(\sigma_X - \rho\sigma_S e^{-\kappa(T-t)})B_t dt + \int_0^T \rho_{IX}^2\sigma_I^2 A(2,2,X_0,S_0,t)dt\Bigg\},\end{aligned}$$

where B_t is given by

$$\begin{aligned}B_t =\ & X_0^2 \exp\left(\sigma_X^2 t - \frac{1}{\kappa}\sigma_S\sigma_X\rho\left(3 - 2e^{-\kappa T} - 2e^{-\kappa t} + e^{-\kappa(T-t)}\right) - S_0\left(e^{-\kappa T} + e^{-\kappa t}\right)\right)\\ \times & \exp\left(\frac{1}{4\kappa}\sigma_S^2\left(2 - e^{-2\kappa T} - e^{-2\kappa t} + 2e^{-\kappa(T-t)} - 2e^{-\kappa(T+t)}\right) - m(2 - e^{-\kappa T} - e^{-\kappa t})\right).\end{aligned}$$

Under the correct model, the 2GBM model, the minimal variance of the realized hedge error is given by

$$\operatorname{Var}(C_T) = c^2 y^2\left(1 - \rho_{IX}^2\right)\left(e^{\sigma_I^2 T} - 1\right).$$

Proof. Since I is a GBM in the 2GBM model, the value function ψ associated with the linear position $h(x) = cx$ is given by $\psi(t,y) = e^{-r(T-t)}cy$. Therefore, $\psi_y(t,y) = e^{-r(T-t)}c$, and the realized error variance in Model 1, following strategy

$$\zeta_t = \rho_{IX}\sigma_I I_t e^{-r(T-t)}c/\left(\sigma_X X_t\right),$$

satisfies

$$C_T = cI_T - e^{rT}\left(v + \int_0^T e^{-rT}c\rho_{IX}\sigma_I I_t dW_t^{(X)}\right),$$

and consequently we have

$$\operatorname{Var}(C_T) = c^2 \operatorname{Var}(I_T) - 2c^2 \operatorname{Cov}\left(I_T, \int_0^T \rho_{IX}\sigma_I I_t dW_t^{(X)}\right) + c^2 E\left(\int_0^T \rho_{IX}^2 \sigma_I^2 I_t^2 dt\right).$$

Note that $\operatorname{Var}(I_T) = A(2,2,X_0,S_0,T) - A^2(1,1,X_0,S_0,T)$ and observe further that

$$E\left(\int_0^T \rho_{IX}^2 \sigma_I^2 I_t^2 dt\right) = \int_0^T \rho_{IX}^2 \sigma_I^2 A(2,2,X_0,S_0,t)dt.$$

It remains to calculate the covariance above. To that effect we recall the decomposition (2.30) of I_T from the proof of Proposition 2.7 and borrow the respective notation to write $\int_0^T \rho_{IX}\sigma_I I_t dW_t^{(X)} = \int_0^T \rho_{IX}\sigma_I I_t d\tilde{W}_t^{(X)} + \int_0^T \rho_{IX}\sigma_I I_t(\sigma_X - \rho\sigma_S e^{-\kappa(T-t)})dt$. Consequently,

$$\begin{aligned}
\operatorname{Cov}\left(I_T, \int_0^T \rho_{IX}\sigma_I I_t dW_t^{(X)}\right) &= X_0 e^{\lambda(T)} E^Q\left(\int_0^T \rho_{IX}\sigma_I I_t(\sigma_X - \rho\sigma_S e^{-\kappa(T-t)})dt\right) \\
&= X_0 e^{\lambda(T)} \int_0^T \rho_{IX}\sigma_I(\sigma_X - \rho\sigma_S e^{-\kappa(T-t)}) E^Q(I_t)\, dt \\
&= \int_0^T \rho_{IX}\sigma_I(\sigma_X - \rho\sigma_S e^{-\kappa(T-t)}) E(I_T I_t)\, dt.
\end{aligned}$$

In order to calculate $E(I_T I_t)$ we proceed similar as in Lemma 2.3. We decompose $I_T I_t$ into

$$\begin{aligned}
I_T I_t = & X_0^2 \exp\left(\int_0^T \left[\sigma_X(1 + 1_{u\le t}) - \sigma_S\rho\left(e^{-\kappa(T-u)} + e^{-\kappa(t-u)}1_{u\le t}\right)\right] dW_u^X\right) \\
&\times \exp\left(-\int_0^T \sigma_S\bar{\rho}\left(e^{-\kappa(T-u)} + e^{-\kappa(t-u)}1_{u\le t}\right) dW_u^{\perp}\right) \\
&\times \exp\left(-\frac{1}{2}\int_0^T \sigma_X^2(1 + 1_{u\le t})du\right) \\
&\times \exp\left(-S_0\left(e^{-\kappa T} + e^{-\kappa t}\right) - m(2 - e^{-\kappa T} - e^{-\kappa t})\right).
\end{aligned}$$

The variances of the stochastic integrals are given by

$$\int_0^T \sigma_X^2(1 + 2_{u\leq t} + 1_{u\leq t}) - 2\sigma_S\sigma_X\rho\left[e^{-\kappa(T-u)} + 1_{u\leq t}\left(2e^{-\kappa(t-u)} + e^{-\kappa(T-u)}\right)\right]$$
$$+ \sigma_S^2\rho^2\left(e^{-\kappa(T-u)} + e^{-\kappa(t-u)}1_{u\leq t}\right)^2 du$$

and

$$\int_0^T \sigma_S^2\bar{\rho}^2\left(e^{-\kappa(T-u)} + e^{-\kappa(t-u)}1_{u\leq t}\right)^2 du.$$

Hence, taking expectation yields

$$E\left(I_T I_t\right) =$$
$$X_0^2 \exp\left(\frac{1}{2}\int_0^T \left[\sigma_X^2 2_{u\leq t} - 2\sigma_S\sigma_X\rho\left[e^{-\kappa(T-u)} + 1_{u\leq t}\left(2e^{-\kappa(t-u)} + e^{-\kappa(T-u)}\right)\right]\right] du\right)$$
$$\times \exp\left(\frac{1}{2}\int_0^T \sigma_S^2\left(e^{-2\kappa(T-u)} + 2e^{-\kappa(T+t-2u)}1_{u\leq t} + e^{-2\kappa(t-u)}1_{u\leq t}\right) du\right)$$
$$\times \exp\left(-S_0\left(e^{-\kappa T} + e^{-\kappa t}\right) - m(2 - e^{-\kappa T} - e^{-\kappa t})\right).$$

The result follows by a simple calculation. □

Proposition 2.5 allows us to analyse the ignorance of a long-term relationship with respect to the variance of the hedge error. Of course the variance of the optimal strategy ξ^* given by Equation (2.13) is less than the standard error using the strategy ζ from the 2GBM model given by Equation (2.24). The upper panels of Figure 2.3 compare the performance of the two strategies.[1] When following the strategy ζ, the risk position is underhedged, yielding the hedge error (dashed line) to grow continually with time to maturity. The hedge error does not flatten as strongly as when following strategy ξ^*, whose corresponding hedge error is depicted by the solid line. For very short maturities, the mean reversion has little time to develop and hence the hedge error entailed by ζ is similar to the hedge error of ξ^*. However, for long maturities ζ is considerably outperformed by ξ^*, leading for example to a more than three times higher error standard deviation over a two year hedging period.

Since by definition there is no strategy with a smaller variance than the variance optimal strategy the proportion of a three times larger value is strongly convincing. To fairly compare our model with the 2GBM model, we also consider the inverse case, that is we study the hedge error of the optimal strategy from the model with the asymptotic

[1] Note that we used the estimated parameters obtained for the August 2009 crude oil futures contract (see Table 2.1) and that the correlation ρ_{IX} can be expressed in terms of the structural parameters of our model, see Equation (2.4).

stationary logspread, ξ^*, when there is no cointegration. To this end we have to derive the resulting hedge error in the 2GBM model which is given in the next proposition.

Proposition 2.6. *Hedging the linear position cI_T with the strategy ξ^*, see Formula* (2.13)*, entails a hedge error in the 2GBM model with variance*

$$\begin{aligned}\operatorname{Var}(C_T) &= c^2\Big\{B(0,2,T)-B^2(0,1,T)\\ &-2\int_0^T \sigma_I\rho\left[1-\frac{\sigma_S}{\sigma_X}\rho e^{-\kappa(T-t)}\right]A(1,1,1,0,T-t)\sigma_X B(1-e^{-\kappa(T-t)},1+e^{-\kappa(T-t)},t)dt\\ &+\int_0^T\left[1-\frac{\sigma_S}{\sigma_X}\rho e^{-\kappa(T-t)}\right]^2 A^2(1,1,1,0,T-t)\sigma_X^2 B(2-2e^{-\kappa(T-t)},2e^{-\kappa(T-t)},t)dt\Big\},\end{aligned} \tag{2.25}$$

where $B(a,b,t)$ is given by

$$B(a,b,t)=E(X_t^aI_t^b)=X_0^aI_0^b\exp\left(\frac{1}{2}t\left[\sigma_X^2(a^2-a)+2ab\sigma_X\sigma_I\rho+\sigma_I^2(b^2-b)\right]\right). \tag{2.26}$$

Proof. Recall from Equation (2.13) that $\xi_t^*=\left[1-\sigma_S\rho e^{-\kappa(T-t)}/\sigma_X\right]\psi_x(t,X_t,S_t)$, with $\psi(t,x,s)=e^{-r(T-t)}E\left(h(X_T^{t,x}e^{-S_T^{t,s}})\right)$. Hence, for $h(y)=cy$ we get $\psi_x(t,x,s)=ce^{-r(T-t)}E\left(X_T^{t,1}e^{-S_T^{t,s}}\right)=ce^{-r(T-t)}A(1,1,1,s,T-t)$. Thus, the realized error variance in the 2GBM model, following the strategy above, satisfies

$$C_T=cI_T-e^{rT}\left(v+c\int_0^T e^{-rT}\left[1-\frac{\sigma_S}{\sigma_X}\rho e^{-\kappa(T-t)}\right]A(1,1,1,S_t,T-t)dX_t\right).$$

Consequently, setting $v=ce^{-rT}E(I_T)$, we have

$$\begin{aligned}\operatorname{Var}(C_T)=&\,c^2\operatorname{Var}(I_T)\\ &-2c^2\operatorname{Cov}\left(I_T,\int_0^T\left[1-\frac{\sigma_S}{\sigma_X}\rho e^{-\kappa(T-t)}\right]A(1,1,1,S_t,T-t)\sigma_X X_t dW_t^{(X)}\right)\\ &+c^2E\left(\int_0^T\left[1-\frac{\sigma_S}{\sigma_X}\rho e^{-\kappa(T-t)}\right]^2A^2(1,1,1,S_t,T-t)\sigma_X^2X_t^2dt\right).\end{aligned}$$

It is straightforward to see that $B(a,b,t)$ defined as $B(a,b,t)=E\left(X_t^aI_t^b\right)$, with X and I as in the 2GBM model, fulfills Equation (2.26). Hence $\operatorname{Var}(I_T)=B(0,2,T)-B^2(0,1,T)$. Observe that we may, via Fubini's theorem, write the expectation in the last term in the variance of C_T above as

$$\int_0^T \left[1 - \frac{\sigma_S}{\sigma_X}\rho e^{-\kappa(T-t)}\right]^2 A^2(1,1,1,0,T-t)\sigma_X^2 E\left(\exp\left(-2S_t e^{-\kappa(T-t)}\right) X_t^2\right) dt.$$

In order to simplify further recall that $S_t = \log(X_t) - \log(I_t)$. Thus,

$$\begin{aligned} E\left(\exp\left(-2S_t e^{-\kappa(T-t)}\right) X_t^2\right) &= E\left(X_t^{2-2e^{-\kappa(T-t)}} I_t^{2e^{-\kappa(T-t)}}\right) \\ &= B(2 - 2e^{-\kappa(T-t)}, 2e^{-\kappa(T-t)}, t), \end{aligned}$$

which combined with the previous integral yields the last term in Expression (2.25). In order to simplify the remaining term in the variance of C_T we apply the standard trick of a change of measure, here with $dQ = (I_T/I_0)dP$, Fubini's Theorem, and reversing the measure change to obtain

$$\int_0^T \sigma_I \rho \left[1 - \frac{\sigma_S}{\sigma_X}\rho e^{-\kappa(T-t)}\right] A(1,1,1,0,T-t)\sigma_X E\left(I_T \exp\left(-S_t e^{-\kappa(T-t)}\right) X_t\right) dt.$$

Finally, using $S_t = \log(X_t) - \log(I_t)$, and the independent increments of a Brownian motion, we get

$$\begin{aligned} &E\left(I_T \exp\left(-S_t e^{-\kappa(T-t)}\right) X_t\right) \\ &\qquad = E\left(I_T I_t^{e^{-\kappa(T-t)}} X_t^{1-e^{-\kappa(T-t)}}\right) \\ &\qquad = E\left(I_t^{1+e^{-\kappa(T-t)}} X_t^{1-e^{-\kappa(T-t)}} \exp\left(\int_t^T \sigma_I dW_u^I - \frac{1}{2}\int_t^T \sigma_I^2 du\right)\right) \\ &\qquad = B(1 - e^{-\kappa(T-t)}, 1 + e^{-\kappa(T-t)}, t), \end{aligned}$$

which together with the previous integral yields the middle term in Expression (2.25) and thus finishes the proof. □

Note that now not all parameters are identified. This is in contrast to the previous scenario, where we investigated the impact of ignoring a long-term relationship. As the logspread is not mean reverting in the 2GBM model the parameter κ is implicitly set to zero. However, to provide a realistic comparison we estimate the implied distribution of $\hat{\kappa}$ in the following way: we simulate 10000 sample paths from the 2GBM model with $T = 885$ observations. Based on these time series we estimate our model leading to the distribution of $\hat{\kappa}$. For the graphical illustration we use the corresponding 10%, 50% and 90% quantiles leading to 0.0927, 0.8365 and 2.2450 respectively. The dashed lines in the lower left panel of Figure 2.3 show the hedge error standard deviation, for these different mean reversion speeds, when the real prices behave as in the 2GBM model, but the risk is hedged according to the cointegration model optimal strategy ξ^*. As a benchmark, the panel depicts also the genuine minimal error standard deviation (solid line) implied

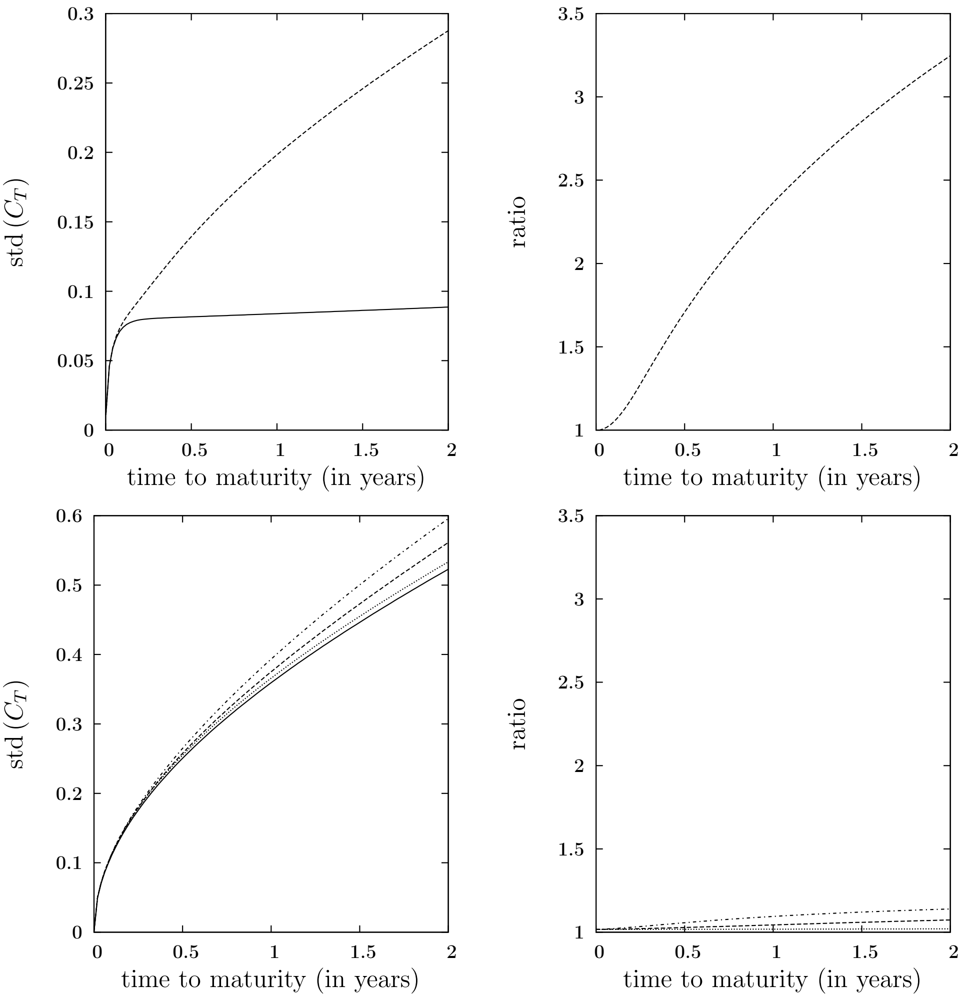

Figure 2.3.: The upper left panel shows the standard deviation of the hedge error under cointegration using the optimal strategy (solid line) and the strategy from the 2GBM model (dashed line). The lower left panel shows the standard deviation under the 2GBM model using the optimal strategy (solid line) and the strategy from the cointegrated model (dashed lines) for $\kappa \in \{2.2450, 0.8365, 0.0927\}$, from top to down, respectively. The panels on the right depict the ratio of the standard deviation of the strategies.

by the optimal strategy ζ. The smaller the mean reversion speed, the smaller the hedge error. Moreover, the hedge error converges to the minimal hedge error as the mean reversion speed converges to zero, showing that the cointegration model embeds the 2GBM model.

In many real world applications, it may not be obvious that there is a long-term relationship between the hedging instrument and the risk to be hedged. Comparing the left upper and left lower panel of Figure 2.3 we conclude that in ambiguous situations it is nevertheless better to use a model allowing for an asymptotic stationary spread rather than using a simpler model that does not. The error implied by mistakenly assuming a mean reverting logspread is significantly smaller than the error made by mistakenly assuming that the hedging instrument is not cointegrated with the risk.

2.4.2. The Costs of Using a Static Hedge

In practice linear positions are often only statically hedged, even though the variance minimizing hedge is not constant. In the numerical example below we compare the standard deviations of static and dynamic hedges of a linear position, and address the question by how much the *dynamic* variance minimizing strategy outperforms the static one.

For that purpose we need to derive the optimal static hedge position $a \in \mathbb{R}$ that minimizes the variance

$$C_T(a) = cI_T - e^{rT}\left(v + \int_0^T e^{-rt} a dX_t\right).$$

With this at hand we can calculate the minimal error standard deviation that can be achieved by hedging statically with futures.

Proposition 2.7. *The optimal static hedging position $\check{a}$ in futures contracts which minimizes the variance of the hedge error is given by*

$$\check{a} = \frac{\operatorname{Cov}\left(cI_T, e^{rT}\int_0^T e^{-rt}dX_t\right)}{\operatorname{Var}\left(e^{rT}\int_0^T e^{-rt}dX_t\right)} = ce^{-rT}\frac{E\left(I_T\int_0^T e^{-rt}dX_t\right)}{\sigma_X^2 E\left(\int_0^T e^{-2rt}X_t^2 dt\right)} \tag{2.27}$$

with corresponding variance

$$\operatorname{Var}\left(C_T(\check{a})\right) = E\left(C_T^2(\check{a})\right) = c^2\operatorname{Var}\left(I_T\right) - \frac{c^2\left[E\left(I_T\int_0^T e^{-rt}dX_t\right)\right]^2}{\sigma_X^2 E\left(\int_0^T e^{-2rt}X_t^2 dt\right)}. \tag{2.28}$$

The expectation in the denominator is given by

$$E\left(\int_0^T e^{-2rt}X_t^2 dt\right) = \begin{cases} \frac{X_0^2}{\sigma_X^2 - 2r}\left(e^{(\sigma_X^2-2r)T} - 1\right), & \text{if } \sigma_X^2 \neq 2r, \\ X_0^2 T, & \text{if } \sigma_X^2 = 2r. \end{cases}$$

For the expectation in the numerator we have, assuming $\sigma_X^2 > r$*,*

$$E\left(I_T \int_0^T e^{-rt} dX_t\right) = \begin{cases} X_0^2 e^{\lambda(T)} \frac{\sigma_X^2}{\sigma_X^2 - r} \left(e^{(\sigma_X^2 - r)T} - 1\right), & \text{if } \rho = 0, \\ X_0^2 e^{\lambda(T) + \frac{\rho\sigma_X\sigma_S}{\kappa} e^{-\kappa T}} (\Lambda_1(T) - \Lambda_2(T)), & \text{if } \rho \neq 0, \end{cases} \tag{2.29}$$

with

$$\begin{aligned} \Lambda_1(T) &= \frac{\sigma_X^2 e^{(\sigma_X^2 - r)T} \left(|\rho|\sigma_X\sigma_S\right)^{-\frac{\sigma_X^2 - r}{\kappa}}}{\kappa^{1 - \frac{\sigma_X^2 - r}{\kappa}}} \left(\gamma\left(\frac{\sigma_X^2 - r}{\kappa}, \frac{1}{\kappa}|\rho|\sigma_X\sigma_S\right)\right. \\ &\qquad \left. -\gamma\left(\frac{\sigma_X^2 - r}{\kappa}, \frac{1}{\kappa}|\rho|\sigma_X\sigma_S e^{-\kappa T}\right)\right), \\ \Lambda_2(T) &= e^{(\sigma_X^2 - r)T} \left(\frac{|\rho|\sigma_X\sigma_S}{\kappa}\right)^{-\frac{\sigma_X^2 - r}{\kappa}} \left(\gamma\left(\frac{\sigma_X^2 - r}{\kappa} + 1, \frac{1}{\kappa}|\rho|\sigma_X\sigma_S\right)\right. \\ &\qquad \left. -\gamma\left(\frac{\sigma_X^2 - r}{\kappa} + 1, \frac{1}{\kappa}|\rho|\sigma_X\sigma_S e^{-\kappa T}\right)\right), \\ \lambda(T) &= -S_0 e^{-\kappa T} - m(1 - e^{-\kappa T}) + \frac{\sigma_S^2}{4\kappa}\left(1 - e^{-2\kappa T}\right) - \frac{\rho\sigma_S\sigma_X}{\kappa}\left(1 - e^{-\kappa T}\right). \end{aligned}$$

Here $\gamma(s, x) = \int_0^x y^{s-1} e^{-y} dy$ *denotes the incomplete Gamma function. Furthermore we have*

$$\mathrm{Var}\,(I_T) = A(2, 2, X_0, S_0, T) - A^2(1, 1, X_0, S_0, T),$$

where A *is as in Equation* (2.17)*.*

Proof. The variance of the hedge error does not depend on the initial capital v, and hence we may assume that $e^{rT} v = cE\,(I_T)$. Holding the constant position a between 0 and T then entails the hedge error

$$C_T(a) = cI_T - E\,(cI_T) - ae^{rT} \int_0^T e^{-rt} dX_t,$$

and since X_t is a martingale we have $E\,(C_T(a)) = 0$. Hence, the variance of $C_T(a)$ is given by

$$
\begin{aligned}
E\left(C_T^2(a)\right) = & c^2 E\left((I_T - E(I_T))^2\right) - 2ace^{rT} E\left((I_T - E(I_T)) \int_0^T e^{-rt} dX_t\right) \\
& + a^2 e^{2rT} E\left(\left(\int_0^T e^{-rt} dX_t\right)^2\right) \\
= & c^2 E\left((I_T - E(I_T))^2\right) - 2ace^{rT} E\left(I_T \int_0^T e^{-rt} dX_t\right) \\
& + a^2 e^{2rT} E\left(\int_0^T e^{-2rt} \sigma_X^2 X_t^2 dt\right).
\end{aligned}
$$

The optimal $\check{a}$ which minimizes the variance of the hedge error is given by

$$
\check{a} = ce^{-rT} \frac{E\left(I_T \int_0^T e^{-rt} dX_t\right)}{\sigma_X^2 E\left(\int_0^T e^{-2rt} X_t^2 dt\right)}.
$$

For the variance of the corresponding hedge error we have

$$
\begin{aligned}
E\left(C_T^2(\check{a})\right) &= \operatorname{Var}(cI_T) - \frac{\operatorname{Cov}\left(cI_T, e^{rT} \int_0^T e^{-rt} dX_t\right)^2}{\operatorname{Var}\left(e^{rT} \int_0^T e^{-rt} dX_t\right)} \\
&= c^2 \operatorname{Var}(I_T) - \frac{c^2 \left[E\left(I_T \int_0^T e^{-rt} dX_t\right)\right]^2}{\sigma_X^2 E\left(\int_0^T e^{-2rt} X_t^2 dt\right)}.
\end{aligned}
$$

The expectation in the denominator is given by

$$
\begin{aligned}
E\left(\int_0^T e^{-2rt} X_t^2 dt\right) &= \int_0^T e^{-2rt} \underbrace{E\left(X_t^2\right)}_{=X_0^2 \exp(\sigma_X^2 t)} dt \\
&= \begin{cases} \frac{X_0^2}{\sigma_X^2 - 2r}\left(e^{(\sigma_X^2 - 2r)T} - 1\right), & \text{if } \sigma_X^2 \neq 2r, \\ X_0^2 T, & \text{if } \sigma_X^2 = 2r. \end{cases}
\end{aligned}
$$

The computations for the expectation in the numerator are somewhat more involved. Using the explicit expressions for S_T and X_T, we may decompose I_T into

$$
I_T = X_T e^{-S_T} = X_0 e^{\lambda(T)} D_T, \tag{2.30}
$$

where D_T is the value at time T of the process D defined by, for all $t \in [0, T]$,

$$
\begin{aligned}
D_t &= \exp\left(\int_0^t (\sigma_X - \rho\sigma_S e^{-\kappa(T-u)}) dW_u^{(X)} - \frac{1}{2}\int_0^t (\sigma_X - \rho\sigma_S e^{-\kappa(T-u)})^2 du \right) \\
&\quad \times \exp\left(- \int_0^t \bar{\rho}\sigma_S e^{-\kappa(T-u)}) dW_u^{\perp} - \frac{1}{2}\int_0^t \bar{\rho}^2\sigma_S^2 e^{-2\kappa(T-u)} du \right)
\end{aligned}
$$

and $\lambda(T)$ is a constant given by

$$
\begin{aligned}
\lambda(T) &= -S_0 e^{-\kappa T} - m(1-e^{-\kappa T}) - \frac{\sigma_X^2}{2}T + \frac{1}{2}\int_0^T (\sigma_X - \rho\sigma_S e^{-\kappa(T-u)})^2 du \\
&\quad + \frac{1}{2}\int_0^T \bar{\rho}^2\sigma_S^2 e^{-2\kappa(T-u)} du \\
&= -S_0 e^{-\kappa T} - m(1-e^{-\kappa T}) + \frac{\sigma_S^2}{4\kappa}(1-e^{-2\kappa T}) - \frac{\rho\sigma_S\sigma_X}{\kappa}(1-e^{-\kappa T}).
\end{aligned}
$$

Note that D is a strictly positive martingale and satisfies Novikov's condition. Therefore we can define a probability measure Q via $dQ = D_T dP$. Under Q the processes $\tilde{W}^{(X)}$ and $\tilde{W}^{\perp}$, for all $t \in [0, T]$,

$$
\begin{aligned}
\tilde{W}_t^{(X)} &= W_t^{(X)} - \int_0^t (\sigma_X - \rho\sigma_S e^{-\kappa(T-u)}) du \\
\tilde{W}_t^{\perp} &= W_t^{\perp} + \int_0^t \bar{\rho}\sigma_S e^{-\kappa(T-u)} du
\end{aligned}
$$

are independent Brownian motions. The dynamics of X, rewritten in terms of $\tilde{W}^{(X)}$, satisfy

$$
dX_t = \sigma_X(\sigma_X - \rho\sigma_S e^{-\kappa(T-t)}) X_t dt + \sigma_X X_t d\tilde{W}_t^{(X)}.
$$

Observe that the expectation of X_t with respect to Q is given by

$$
\begin{aligned}
E^Q(X_t) &= X_0 \exp\left(\int_0^t \sigma_X(\sigma_X - \rho\sigma_S e^{-\kappa(T-u)}) du \right) \\
&= X_0 \exp\left(\sigma_X^2 t + \frac{\rho\sigma_X\sigma_S}{\kappa} e^{-\kappa T} \right) \exp\left(-\frac{\rho\sigma_X\sigma_S}{\kappa} e^{-\kappa(T-t)} \right).
\end{aligned}
$$

Now the expectation term in the numerator can be written as

$$\begin{aligned} E\left(I_T \int_0^T e^{-rt} dX_t\right) &= X_0 e^{\lambda(T)} E^Q\left(\int_0^T e^{-rt} dX_t\right) \qquad (2.31) \\ &= X_0 e^{\lambda(T)} \int_0^T e^{-rt}\sigma_X(\sigma_X - \rho\sigma_S e^{-\kappa(T-t)}) E^Q\left(X_t\right) dt \\ &= X_0^2 e^{\lambda(T) + \frac{\rho\sigma_X\sigma_S}{\kappa} e^{-\kappa T}} \\ &\quad \times \underbrace{\int_0^T e^{(\sigma_X^2 - r)t}(\sigma_X^2 - \rho\sigma_X\sigma_S e^{-\kappa(T-t)}) e^{-\frac{\rho\sigma_X\sigma_S}{\kappa} e^{-\kappa(T-t)}} dt}_{=A}\,. \end{aligned}$$

For $\rho = 0$ we are done. For $\rho \neq 0$ we continue by substituting $u = |\rho|\sigma_X\sigma_S e^{-\kappa(T-t)}/\kappa$ which leads to an explicit expression for the above integral A in terms of the incomplete Gamma function $\gamma(s,x) = \int_0^x y^{s-1}e^{-y}dy$.

$$\begin{aligned} A &= e^{(\sigma_X^2 - r)T}\left(\frac{|\rho|\sigma_X\sigma_S}{\kappa}\right)^{-\frac{\sigma_X^2 - r}{\kappa}} \int_{\frac{1}{\kappa}|\rho|\sigma_X\sigma_S e^{-\kappa T}}^{\frac{1}{\kappa}|\rho|\sigma_X\sigma_S} u^{\frac{\sigma_X^2 - r}{\kappa}} (\sigma_X^2 - \kappa u) e^{-u} \frac{1}{\kappa u} du \\ &= \frac{\sigma_X^2 e^{(\sigma_X^2 - r)T}\left(|\rho|\sigma_X\sigma_S\right)^{-\frac{\sigma_X^2 - r}{\kappa}}}{\kappa^{1 - \frac{\sigma_X^2 - r}{\kappa}}} \int_{\frac{1}{\kappa}|\rho|\sigma_X\sigma_S e^{-\kappa T}}^{\frac{1}{\kappa}|\rho|\sigma_X\sigma_S} u^{\frac{\sigma_X^2 - r}{\kappa} - 1} e^{-u} du \\ &\quad - e^{(\sigma_X^2 - r)T}\left(\frac{|\rho|\sigma_X\sigma_S}{\kappa}\right)^{-\frac{\sigma_X^2 - r}{\kappa}} \int_{\frac{1}{\kappa}|\rho|\sigma_X\sigma_S e^{-\kappa T}}^{\frac{1}{\kappa}|\rho|\sigma_X\sigma_S} u^{\frac{\sigma_X^2 - r}{\kappa}} e^{-u} du \\ &= \Lambda_1(T) - \Lambda_2(T), \end{aligned}$$

where

$$\begin{aligned} \Lambda_1(T) = \frac{\sigma_X^2 e^{(\sigma_X^2 - r)T}\left(|\rho|\sigma_X\sigma_S\right)^{-\frac{\sigma_X^2 - r}{\kappa}}}{\kappa^{1 - \frac{\sigma_X^2 - r}{\kappa}}} &\left(\gamma\left(\frac{\sigma_X^2 - r}{\kappa}, \frac{1}{\kappa}|\rho|\sigma_X\sigma_S\right)\right. \\ &\left. - \gamma\left(\frac{\sigma_X^2 - r}{\kappa}, \frac{1}{\kappa}|\rho|\sigma_X\sigma_S e^{-\kappa T}\right)\right) \end{aligned}$$

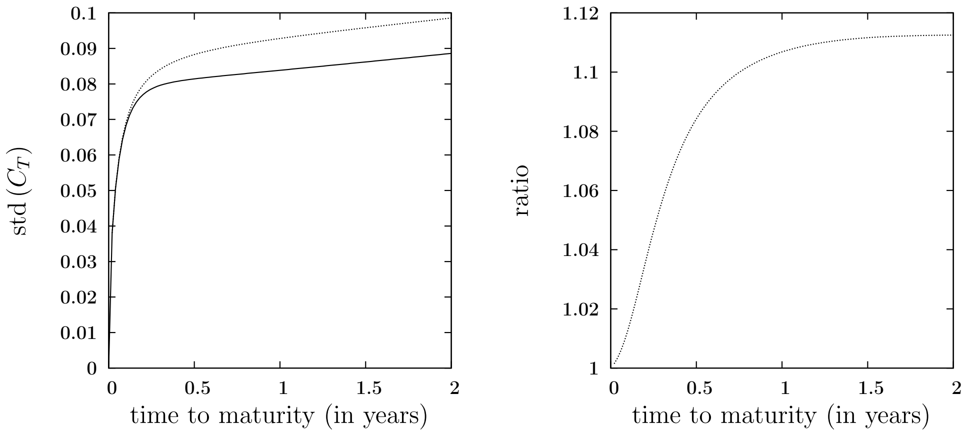

Figure 2.4.: The panel on the left shows the standard deviation of the static (dashed line) versus the dynamic hedge error (solid line). The right hand panel depicts the ratio. As we have to assume a constant interest rate for the computation of the variance of the static hedge error we fix it at $r = 0.02$.

and

$$\Lambda_2(T) = e^{(\sigma_X^2 - r)T} \left(\frac{|\rho|\sigma_X\sigma_S}{\kappa}\right)^{-\frac{\sigma_X^2 - r}{\kappa}} \left(\gamma\left(\frac{\sigma_X^2 - r}{\kappa} + 1, \frac{|\rho|\sigma_X\sigma_S}{\kappa}\right) - \gamma\left(\frac{\sigma_X^2 - r}{\kappa} + 1, \frac{|\rho|\sigma_X\sigma_S e^{-\kappa T}}{\kappa}\right)\right).$$

Plugging this into Equation (2.31) gives

$$E\left(I_T \int_0^T e^{-rt} dX_t\right) = X_0^2 e^{\lambda(T) + \frac{\rho\sigma_X\sigma_S}{\kappa} e^{-\kappa T}} (\Lambda_1(T) - \Lambda_2(T)).$$

The expression for the Var (I_T) is straightforward. □

Remark 2.8. Note that the assumption $\sigma_X^2 > r$ in the Proposition above is only needed in order to get a closed expression with respect to the incomplete Gamma function. In case $\sigma_X^2 \le r$ the defining integral of the incomplete Gamma function explodes around 0. In this case Equation (2.29) still holds if we replace γ with the *upper incomplete Gamma function* $\gamma(s,x) = -\int_x^\infty y^{s-1} e^{-y} dy$. In any case, regardless of $\sigma_X^2 > r$, Formula (2.29) holds when we replace $\Lambda_1(T) - \Lambda_2(T)$ with $\int_0^T e^{(\sigma_X^2 - r)t}(\sigma_X^2 - \rho\sigma_X\sigma_S e^{-\kappa(T-t)}) e^{-\rho\sigma_X\sigma_S e^{-\kappa(T-t)}/\kappa} dt$. ♦

Proposition 2.9. *The expressions for the optimal static hedge* $\check{a}$ (2.27) *and for the corresponding variance* (2.28) *from the previous proposition hold also in the* $2GBM$

case. For the involved expectations we have

$$E\left(\int_0^T e^{-2rt} X_t^2 dt\right) = \begin{cases} \frac{X_0^2}{\sigma_X^2 - 2r}(e^{(\sigma_X^2-2r)T} - 1), & \textit{if } \sigma_X^2 \neq 2r, \\ X_0^2 T, & \textit{if } \sigma_X^2 = 2r, \end{cases}$$

and

$$E\left(I_T \int_0^T e^{-rt} dX_t\right) = \begin{cases} X_0 I_0 \frac{\rho_{IX}\sigma_X\sigma_I}{\rho_{IX}\sigma_X\sigma_I - r}(e^{(\rho_{IX}\sigma_X\sigma_I - r)T} - 1), & \textit{if } \rho_{IX}\sigma_X\sigma_I \neq r, \\ X_0 I_0 \rho_{IX}\sigma_X\sigma_I T, & \textit{if } \rho_{IX}\sigma_X\sigma_I = r. \end{cases}$$

Furthermore

$$\mathrm{Var}\,(I_T) = I_0^2(e^{\sigma_I^2 T} - 1).$$

Proof. The expressions for $\check{a}$ and the corresponding variance $\mathrm{Var}\,(C_T(\check{a}))$ are model independent and therefore the same as in Proposition 2.7. In both models X is a GBM and hence we again have

$$E\left(\int_0^T e^{-2rt} X_t^2 dt\right) = \begin{cases} \frac{X_0^2}{\sigma_X^2 - 2r}(e^{(\sigma_X^2-2r)T} - 1), & \text{if } \sigma_X^2 \neq 2r, \\ X_0^2 T, & \text{if } \sigma_X^2 = 2r. \end{cases}$$

For the expectation in the numerator we get

$$E\left(I_T \int_0^T e^{-rt} dX_t\right) = E\left(\int_0^T e^{-rt} d\langle I, X\rangle_t\right) = \int_0^T e^{-rt} \rho_{IX}\sigma_X\sigma_I E\,(I_t X_t)\,dt.$$

Straightforward computations give $E\,(X_t I_t) = X_0 I_0 e^{\rho_{IX}\sigma_X\sigma_I t}$, which plugged into the above expressions gives the desired result. The expression for the $\mathrm{Var}\,(I_T)$ is straightforward. □

Using the expressions for the standard deviation of the hedge error from Theorem 2.4 and Proposition 2.7 we can compare the risks entailed by both strategies. The left hand panel of Figure 2.4 depicts the standard deviation of the static and dynamic variance minimizing hedge against time to maturity. The right hand panel shows the increase of the standard deviation if one confines with the static hedge. The increase in the variability by more than 10% for positions hedged over a period of one year indicates that the hedge should be dynamically adjusted. The figure is again based on the estimated parameter values obtained for the August 2009 crude oil futures contract (see Table 2.1 in Section 2.1.2).

2.5. Including Directional Views and Stochastic Volatility

In this section we extend the model introduced in Section 2.1 by allowing for stochastic volatility of the futures and the logspread. We assume that the volatility of both processes are proportional to a Cox-Ingersoll-Ross process. The futures dynamics thus coincides with the dynamics of the risky asset in the Heston model.

Let (W^1, W^2, W^3) be a 3-dimensional Brownian motion and suppose that the futures price process X and its volatility $\nu = (\nu_t)_{t \geq 0}$ evolve according to the SDE

$$\begin{aligned} dX_t &= \mu(t, \nu_t) X_t dt + \sqrt{\nu_t} X_t dW_t^1 \\ d\nu_t &= \beta(\vartheta - \nu_t) dt + \sigma_\nu \sqrt{\nu_t} (\rho_1 dW_t^1 + \bar{\rho}_1 dW_t^2), \end{aligned}$$

where $\rho_1 \in [-1, 1]$, $\bar{\rho}_1 = \sqrt{1 - \rho_1^2}$, β, ϑ, $\sigma_\nu > 0$, and $\mu : \mathbb{R}_+^2 \to \mathbb{R}$ is measurable. As before, let $S_t = \log(X_t) - \log(I_t)$. Assume that the logspread's volatility is proportional to ν, and that S solves the mean reverting SDE

$$dS_t = \kappa(m - S_t) dt + \sigma_S \sqrt{\nu_t} (\rho dW_t^1 + \bar{\rho}\eta dW_t^2 + \bar{\rho}\bar{\eta} dW_t^3), \quad S_0 = s.$$

Since we have included a directional view in the dynamics of the futures price, we cannot directly invoke the method used in Section 2.2 for the derivation of the variance optimal hedge. When the trading instruments are assumed to be trended, and hence are not martingales, then it is very difficult to determine variance optimal hedging strategy. There are, however, other quadratic optimality criteria that considerably simplify the calculation of hedging strategies. A very intriguing type of hedging strategies are the so-called *locally risk minimizing hedging strategies*. These are variance optimal strategies with respect to a particular martingale measure, usually referred to as the *minimal martingale measure*. For an overview on quadratic hedging approaches we refer to Schweizer [2001] and for some more details on local risk minimization to Section 3.1.

In our extended model the minimal martingale measure $\widehat{Q}$ is given by

$$\frac{d\widehat{Q}}{dP} = \exp\left(-\int_0^T \omega(t, \nu) dW_t^1 - \frac{1}{2} \int_0^T \omega^2(t, \nu) dt \right),$$

where $\omega(t, \nu_t) = \mu(t, \nu_t)/\sqrt{\nu_t}$ is the market price of risk.[2]

One can proceed as in Section 2.2 for the derivation of the local risk minimizing hedge, i.e. the variance optimal hedge relative to $\widehat{Q}$. The *value function* of $h(I_T)$ will be defined by

$$\psi(t, x, v, s) = e^{-r(T-t)} E^{\widehat{Q}} \left(h(X_T^{t,x,v} e^{-S_T^{t,s}}) \right).$$

With the same assumptions on h one can show that the local risk minimizing hedge is given by

$$\widehat{\xi}_t = \psi_x(t, X_t, \nu_t, S_t) \left[1 - \sigma_S \rho e^{-\kappa(T-t)} \right] + \frac{\sigma_\nu \rho_1}{X_t} \psi_v(t, X_t, \nu_t, S_t).$$

Observe that the local risk minimizing hedge is now a weighted sum of the Delta

[2] A sufficient condition for this to be a proper measure change is the following growth condition on ω. For $A, B \geq 0$ and $\delta \in [0, 1/2]$ we assume $|\omega(t, x)| \leq A + Bx^\delta$, $x \geq 0$.

and Vega of the risk position's expectation under the minimal martingale measure $\widehat{Q}$. Clearly, the term involving the Vega of the position appears due to the additional non-tradable risk induced by the stochastic volatility, which also needs to be cross-hedged.

A similar analysis as in the previous sections, e.g. estimation of model parameters, derivation of hedge errors and their respective standard deviations, is somewhat more involved. However, one can profit of the affine model structure and express the value function and its gradient in terms of generalized Ricatti equations. Fourier inversion methods then yield semi-explicit formulas for optimal strategies, which are amenable to swift simulation analysis.

2.6. Conclusion and Outlook

Hedging is essential for controlling and managing risk and it is an important area of research. In this chapter we show that a long-term relationship between the risk and the hedging instrument has important implications for the optimal hedging strategy and, thus, also for the hedge error. In particular, we propose a model which explicitly takes into account such a long-term relationship. We derive the variance optimal cross-hedge strategy und provide the variance of the hedge error in terms of the model's parameters. We demonstrate the practical relevance of incorporating the long-term relationship through an empirical example, where we find a long-term relationship between most crude oil futures contracts and the spot kerosene price. Interestingly, the model is also consistent with the commonly observed behavior of commodity traders, who use for cross hedges a hedge ratio of one instead of a hedge ratio dampened by the cross correlation between the risk and the hedging instrument, which is implied by models with only correlated Brownian motions. Furthermore, we show that even for cases where the decision concerning the asymptotic stationarity of the logspread is not obvious, it is better to allow for a long-term relationship rather than to neglect it.

The model can be extended towards several directions to provide a more realistic dynamics for asset prices. Especially the consideration of jumps in the price process seems to be an interesting extension. However, several specifications are plausible and a careful empirical investigation is needed. On an ad hoc basis it is, for example, not clear whether the price processes jump together and how the jump sizes are related. These aspects will have significant impact on the properties of the hedge error and we plan to investigate these questions in more detail in future research.

3. Stochastic correlation

In this chapter we extent contemporary results on quadratic hedging with basis risk by allowing for the correlation to be random. As usual, we assume that the price of the tradable asset and the value of the non-tradable index evolve according to geometric Brownian motions. However, we will assume that the correlation between the driving Brownian motions is not constant, but a random process with values between -1 and 1. More precisely, we will assume that the correlation process is the solution of a stochastic differential equation. Then, our focus is on deriving an explicit representation of the locally risk minimizing strategy in terms of simple expectations.

The chapter is organised as follows. In Section 3.1 we give a short introduction into local risk minimization. Section 3.2 introduces our model and gives an overview on the main results we obtained. The details we use to derive our hedge formula are provided in Section 3.3. We continue in Section 3.4 by analyzing the boundary behaviour and integrability properties of correlation processes. We conclude with Section 3.5 by giving some explicit examples of correlation processes for which our main results hold.

3.1. A brief review of local risk minimization

In this section we give a short introduction into the theory of local risk minimization in a quadratic sense. The material presented is a streamlined version of Schweizer [2008].

We start with a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, P)$, where $T > 0$ is a finite time horizon and the filtration $(\mathcal{F}_t)$ satisfies the usual conditions, that is $(\mathcal{F}_t)$ is right continuous and completed by the P-null sets. We consider a financial market with one risky asset S and one non-risky asset, say a money market account with dynamics B. We suppose that the discounted asset price $X = S/B$ is an $\mathbb{R}$-valued continuous semimartingale, and we assume that X satisfies the so-called *structure condition (SC)*. This means that X is a special semimartingale with canonical decomposition

$$X = X_0 + M + A = X_0 + M + \int \lambda d\langle M\rangle,$$

where M is a locally square integrable martingale with $M_0 = 0$, and λ is an $\mathbb{R}$-valued, predictable process such that the *mean-variance tradeoff* process $K = \int \lambda dA = \int \lambda^2 d\langle M\rangle$ satisfies $K_T < \infty$, P-almost surely. It is well known that (SC) is related to an absence-of-arbitrage condition; see Schweizer [2008] for a reference.

Definition 3.1. (Schweizer [2008, Definition 1.1])
The space Θ_S consists of all $\mathbb{R}$-valued predictable processes ξ such that the stochastic

integral process $\int \xi dX$ is well-defined and

$$E\left[\int_0^T \xi_s^2 d\langle M\rangle_s + \left(\int_0^T |\xi_s dA_s|\right)^2\right] < \infty.$$

An L^2-*strategy* is a pair $\varphi = (\xi, \eta)$, where $\xi \in \Theta_S$ and η is a real-valued adapted process such that the *value process* $V(\varphi) = \xi X + \eta$ is right-continuous and square-integrable. φ is called *0-achieving* if $V_T(\varphi) = 0$, P-almost surely.

As usual, a strategy $\varphi = (\xi, \eta)$ describes the investment decisions of an agent trading in the financial market. An investor following the strategy φ holds ξ_t shares of the discounted asset X at time t, and keeps η_t units in a money market account. In this section we use the money market account as numeraire so that we need not to bother about the interest rate.

We next consider a payment stream $H = (H_t)_{0\le t\le T}$ kept fixed throughout this introduction. Mathematically, H is right-continuous, adapted, real-valued and square-integrable; the interpretation is that $H_t \in L^2(P)$ represents the total payments on $[0, t]$ arising due to some financial contract. A European contingent claim with maturity T would have $H_t = 0$, for all $t < T$, and just an $\mathcal{F}_T$-measurable payoff $H_T \in L^2(P)$ due at time T; in general, the process H involves both cash inflows and outlays, and can but need not be of finite variation.

Definition 3.2. (Schweizer [2008, Definition 1.2])
Fix a payment stream H. The *(cumulative) cost process* of an L^2-strategy $\varphi = (\xi, \eta)$ is

$$C_t^H(\varphi) = H_t + V_t(\varphi) - \int_0^t \xi_s dX_s, \qquad 0 \le t \le T.$$

φ is called *self-financing* (for H) if $C_t^H(\varphi)$ is constant, and *mean-self-financing* if $C_t^H(\varphi)$ is a martingale (which is then square-integrable). Under the assumption that X fulfills (SC) and that the mean-variance tradeoff process K is continuous we say that an L^2-strategy φ is *locally risk minimizing* if φ is 0-achieving and mean-self-financing, and the cost process $C^H(\varphi)$ is strongly orthogonal to M, that is $\langle M, C^H(\varphi)\rangle_t = 0$, for all $t \in [0, T]$.

Thus, $C_t(\varphi)$ comprises the hedger's accumulated costs during $[0, t]$ including the payments H_t, and $V_t(\varphi)$ should therefore be interpreted as the value of the portfolio $\varphi_t = (\xi_t, \eta_t)$ held at time t after the payments H_t. In particular, $V_T(\varphi)$ is the value of the portfolio φ_T upon settlement of all liabilities, and a natural condition is then to restrict to 0-achieving strategies as defined in Definition 3.1.

Remark 3.3. (Schweizer [2008, Remark 1.3])
Observe that if $\varphi_t = (\xi_t, \eta_t)$ is a 0-achieving and mean-self-financing L^2-strategy for H, then φ is uniquely determined from ξ (and of course H). ♦

It is well known that the locally risk-minimizing strategy can be obtained via the so-called *Föllmer-Schweizer decomposition (FS)* of the final payment H_T. That is the decomposition of H_T into

$$H_T = H_T^{(0)} + \int\limits_0^T \xi_s^{H_T} dX_s + L_T^{H_T}, \qquad P\text{-almost surely,} \tag{3.1}$$

where $H_T^{(0)} \in L^2(P)$ is $\mathcal{F}_0$-measurable, ξ^{H_T} is in Θ_S, and the process L^{H_T} is a (right-continuous) square-integrable martingale strongly orthogonal to M and satisfying $L_0^{H_T} = 0$. Notice that such a decomposition can be shown to be unique. Once we have (3.1), the desired strategy $\varphi = (\xi, \eta)$ is then given by

$$\xi = \xi^{H_T}, \qquad \eta = V^{H_T} - \xi^{H_T} X,$$

with

$$V_t^{H_T} = H_T^{(0)} + \int\limits_0^t \xi_s^{H_T} dX_s + L_t^{H_T} - H_t, \qquad 0 \le t \le T,$$

see Schweizer [2008, Proposition 5.2]. Furthermore, the associated cost process is given by

$$C_t^H(\varphi) = H_T^{(0)} + L_t^{H_T}, \qquad 0 \le t \le T.$$

3.2. The model and the main results

Let $W = (W^1, W^2, W^3)$ be a three-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \le t \le T}, P)$, where the filtration is generated by W and augmented by the P-null sets. Consider two processes with dynamics

$$\begin{aligned} dS_t &= S_t \left(\mu_X dt + \sigma_X dW_t^1\right), \\ dU_t &= U_t \left(\mu_I dt + \sigma_I \left(\rho_t dW_t^1 + \sqrt{1-\rho_t^2} dW_t^2\right)\right), \end{aligned}$$

where W^1 and W^2 are independent Brownian motions. To simplify the presentation we will assume throughout that all coefficients are constant, more precisely $\mu_X, \mu_I \in \mathbb{R}$ and $\sigma_X, \sigma_I \in \mathbb{R} \setminus \{0\}$.

We assume that S is the price process of a tradable asset, and U the value process of a non-tradable index. The correlation ρ is assumed to follow

$$d\rho_t = a(\rho_t)dt + g(\rho_t)d\hat{W}_t, \tag{3.2}$$

where $\hat{W}$ is given by $\hat{W} = \gamma W_u^1 + \delta W_u^2 + \sqrt{1-\gamma^2-\delta^2} W_u^3$, with W^3 being a Brownian motion independent of W^1 and W^2, and γ and δ real numbers such that $\delta^2 + \gamma^2 \le 1$.

For the moment, we assume that the coefficients of the correlation dynamics, a and g, belong to $\mathcal{C}^1(-1,1)$, and that there exists a unique solution ρ of (3.2) with values in $[-1,1]$.

Throughout we suppose that the interest $r>0$ is constant, and let $B_t = e^{rt}$. The discounted processes S and U will be denoted by

$$X_t = e^{-rt}S_t, \qquad I_t = e^{-rt}U_t.$$

Notice that

$$\begin{aligned} dX_t &= X_t\left(\left(\mu_X - r\right)dt + \sigma_X dW_t^1\right), \\ dI_t &= I_t\left(\left(\mu_I - r\right)dt + \sigma_I\left(\rho_t dW_t^1 + \sqrt{1-\rho_t^2}dW_t^2\right)\right). \end{aligned} \tag{3.3}$$

Consider a derivative $d(U_T)$ depending on the non-tradable index. Let $h(x) = e^{-rT}d(e^{rT}x)$. Then $d(U_T) = e^{rT}h(I_T)$. Our goal is to analyse how to hedge the liability $h(I_T)$ by trading the asset X.

Since the market is incomplete we need to choose a criteria according to which strategies are chosen and the prices of contingent claims are computed. We will use the framework of local risk minimization of Section 3.1.

Our first main result is an explicit hedge formula, which can be easily implemented, for example by simple Monte Carlo simulation. We will state it right away in Theorem 3.4, after a brief collection of some notations and assumptions.

We will need the conditional versions of the processes I and ρ, which are given by

$$I_s^{t,y,v} = y + \int_t^s I_u^{t,y,v}\left(\left(\mu_I - r\right)du + \sigma_I\left(\rho_u^{t,v}dW_u^1 + \sqrt{1-\left(\rho_u^{t,v}\right)^2}dW_u^2\right)\right), \tag{3.4}$$

$$\rho_s^{t,v} = v + \int_t^s a\left(\rho_u^{t,v}\right)du + \int_t^s g\left(\rho_u^{t,v}\right)d\hat{W}_u, \tag{3.5}$$

for $t\in[0,T)$, $(y,v)\in\mathbb{R}^+\times(-1,1)$.

In order to find a nice representation of the quadratic hedge, we also need the dynamics of the derivatives of $I^{t,y,v}$ and $\rho^{t,v}$ with respect to the initial values y and v. Note that the derivative with respect to y of $I^{t,y,v}$ is given by $\frac{\partial}{\partial y}I^{t,y,v} = \frac{I^{t,y,v}}{y} = I^{t,1,v}$, and obviously $\frac{\partial}{\partial y}\rho^{t,v} = 0$. If the correlation process neither attains -1 nor 1 up to time T, then the derivatives of $I^{t,y,v}$ and $\rho^{t,v}$ with respect to v are defined. Moreover, the processes $\frac{\partial}{\partial v} I^{t,y,v}$ and $\frac{\partial}{\partial v}\rho^{t,v}$, denoted by $\bar{I}^{t,y,v}$ and $\bar{\rho}^{t,v}$ respectively, solve the SDE

$$\bar{I}_s^{t,y,v} = \int_t^s \bar{I}_u^{t,y,v}\left((\mu_I - r)\,du + \sigma_I\left(\rho_u^{t,v} dW_u^1 + \sqrt{1-\left(\rho_u^{t,v}\right)^2} dW_u^2\right)\right)$$

$$+ \int_t^s I_u^{t,y,v}\sigma_I\left(\bar{\rho}_u^{t,v} dW_u^1 - \frac{\rho_u^{t,v}}{\sqrt{1-\left(\rho_u^{t,v}\right)^2}}\bar{\rho}_u^{t,v} dW_u^2\right), \tag{3.6}$$

$$\bar{\rho}_s^{t,v} = 1 + \int_t^s a'\left(\rho_u^{t,v}\right)\bar{\rho}_u^{t,v} du + \int_t^s g'\left(\rho_u^{t,v}\right)\bar{\rho}_u^{t,v} d\hat{W}_u, \tag{3.7}$$

for $s \in [t,T]$, see Protter [2004, Chapter V.7, Theorem 38]. Notice that the correlation boundaries -1 and 1 are not attained if and only if the stopping times $\tau^v = \tau^{t,v} = \inf\{s \geq t : \rho_s^{t,v} \in \{-1,1\}\}$ satisfy $\tau^{t,v} > T$, P-almost surely. We formulate this as Condition

(H1) $\quad \tau^{t,v} > T, \qquad P$-almost surely, $\forall\, v \in (-1,1)$.

Notice that (H1) guarantees that $\int_t^T (g'(\rho_u^{t,v}))^2 du < \infty$, P-almost surely and

$$\int_t^T \frac{(\bar{\rho}_u^{t,v})^2}{1-\left(\rho_u^{t,v}\right)^2} du < \infty, \qquad P\text{-almost surely}, \tag{3.8}$$

and hence the stochastic integrals appearing in (3.6) and (3.7) are defined. Finally, for our aim of deriving an explicit representation of the quadratic hedge, we need to impose a stronger integrability condition than (3.8) on ρ and $\bar{\rho}$. More precisely, we will assume that the following condition is satisfied

(H2) There exists a $p > 1$ such that for every $t \in [0,T]$ and $v_0 \in (-1,1)$ there exists an open neighbourhood $U \subset (-1,1)$ of v_0 such that

$$\sup_{v\in U} E \int_t^T \left|\frac{(\bar{\rho}_s^{t,v})^2}{1-\left(\rho_s^{t,v}\right)^2}\right|^p ds < \infty.$$

We are now ready to state our first main result which gives an explicit expression for the locally risk minimizing strategy in terms of expectations with respect to the measure $\tilde{P}$ with density $d\tilde{P}/dP = \mathcal{E}(-\int \frac{\mu_X - r}{\sigma_X} dW^1)_T$, where $\mathcal{E}(\cdot)$ denotes the Doléans-Dade exponential. Note that $\tilde{P}$ corresponds to the so-called *minimal martingale measure*, see Schweizer [2008].

Theorem 3.4. *Suppose that the coefficients a and g in the dynamics of ρ are continuously differentiable on $(-1,1)$. Assume furthermore that (H1) and (H2) are satisfied. Let h be Lipschitz such that the weak derivative h' is Lebesgue-almost everywhere continuous. Then, the locally risk minimizing strategy $\varphi = (\xi,\eta)$ exists for the derivative*

$h(I_T)$. ξ is given by $\xi_t = \xi(t, I_t, X_t, \rho_t)$, for $t \in [0,T]$, where

$$\xi(t,y,x,v) = \frac{y}{x}\left[v\frac{\sigma_I}{\sigma_X}\tilde{E}\left[h'\left(I_T^{t,y,v}\right)I_T^{t,1,v}\right] + \frac{g(v)\gamma}{\sigma_X}\tilde{E}\left[h'\left(I_T^{t,y,v}\right)\bar{I}_T^{t,1,v}\right]\right]. \tag{3.9}$$

The proof of this Theorem is postponed to Section 3.3.

Remark 3.5. In terms of the original processes S and U the hedge would be given by $\xi_t = \hat{\xi}(t, U_t, S_t, \rho_t)$ where

$$\hat{\xi}(t,y,x,v) = \frac{y}{x}\left[v\frac{\sigma_I}{\sigma_X}\tilde{E}\left[d'\left(U_T^{t,y,v}\right)U_T^{t,1,v}\right] + \frac{g(v)\gamma}{\sigma_X}\tilde{E}\left[d'\left(U_T^{t,y,v}\right)\bar{U}_T^{t,1,v}\right]\right]e^{-r(T-t)},$$

with $\bar{U}$ obtained in the same way as $\bar{I}$. ♦

Before we state our second main contribution, let us apply the previous result to derive the locally risk minimizing strategy for a European Call option.

Corollary 3.6. *Suppose that the correlation is a deterministic function of time. For strike $K > 0$, let $d(x) = \max\{(x-K), 0\}$. Then, the locally risk minimizing strategy $\varphi = (\xi, \eta)$ exists for the derivative $d(U_T)$. ξ is given by*

$$\xi_t = \rho_t \frac{U_t \sigma_I}{S_t \sigma_X}\Delta_{BS}(t, U_t, \kappa_t, \sigma_I),$$

for $t \in [0,T]$, where $\kappa_t = -(\mu_I - r)(T-t) + \sigma_I(\frac{\mu_X - r}{\sigma_X})\int_t^T \rho_s ds$ and, with Φ the standard normal cumulative distribution function,

$$\Delta_{BS}(t,y,q,\sigma) = \exp(-q)\Phi\left(\frac{\ln(y/K) + \left(r + \sigma_I^2/2\right)(T-t) - q}{\sigma_I\sqrt{T-t}}\right)$$

is the Black-Scholes delta for options on stocks with continuous dividend yield q.

The content of the preceding corollary is only a slight extension of a result already obtained in Hulley and McWalter [2008] for the case of constant correlation. The proof is a simple straightforward calculation.

Remark 3.7. From the local risk minimizing strategy we can easily deduce the so-called *mean-variance* optimal hedging strategy for the payoff $h(I_T)$. The mean variance hedge is defined to be the self-financing strategy minimzing the variance of the *global* hedging error, and usually differs from the local risk minimizing strategy, see Schweizer [2001] for an introduction into mean-variance hedging. In the model considered here, the mean variance trade-off process is deterministic, and hence, by an appeal to Schweizer [2001, Theorems 4.6 and 4.7], the mean variance hedge has a representation allowing to derive it numerically by a simply recursion. Namely, letting $w = \tilde{E}(h(I_T))$, the mean-variance optimal strategy $(\tilde{\xi}, \tilde{\eta})$ for $h(I_T)$ satisfies, with $\tilde{\xi} = \xi^{(w)}$,

$$\xi_t^{(w)} = \xi_t + \frac{\mu_X - r}{\sigma_X^2 X_t}\left[\tilde{E}\left[h\left(I_T\right)|\mathcal{F}_t\right] - w - \int_0^t \xi_s^{(w)} dX_s\right],$$

for all $t \in [0,T]$, and

$$\tilde{\eta}_t = w + \int_0^t \xi_s^{(w)} dX_s - \xi_t^{(w)} X_t,$$

for all $t \in [0,T]$. ♦

Our second main result concerns conditions on the coefficients a and g of ρ such that (H1) and (H2) are fulfilled.

Theorem 3.8. *Let a and g be continuously differentiable with bounded derivatives. We assume that $g(-1) = g(1) = 0$, and that g does not have any roots in $(-1,1)$. If*

$$\limsup_{x \uparrow 1} \frac{2a(x)(1-x)}{g^2(x)} < 0 \text{ and } \liminf_{x \downarrow -1} \frac{2a(x)(1+x)}{g^2(x)} > 0, \tag{3.10}$$

then (H1) and (H2) are satisfied, and hence, the delta hedge is given as in Theorem 3.4.

The preceding theorem can be generalized, which will enable us to give an example in Section 3.5 where the derivative of g is unbounded. This, however, requires a little more notation, which for ease of exposition is avoided here. See Section 3.4 for a proof of Theorem 3.8 and the more general Proposition 3.20.

3.3. Derivation of the hedge formula

In this section we will derive and prove the hedge formula stated in Theorem 3.4. In Subsection 3.3.1 we use BSDEs to derive the Föllmer-Schweizer decomposition, which is the key to obtain the Formula (3.9). In Sections 3.3.2 and 3.3.3 we provide details which we need along the way. It is there that we need Conditions (H1) and (H2). Let us first recall the definition of a BSDE.

As in Section 3.2, let $W = (W^1, W^2, W^3)$ be a three-dimensional Brownian motion. The filtration generated by W and completed by the P-null sets will be denoted by $(\mathcal{F}_t)$. Let $T > 0$ and ξ be an $\mathcal{F}_T$-measurable random variable, and let $f : \Omega \times [0,T] \times \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}$ be a measurable function such that for all $(y,z) \in \mathbb{R} \times \mathbb{R}^3$ the mapping $f(\cdot,\cdot,y,z)$ is predictable. A solution of the BSDE with *terminal condition* ξ and *generator* f is defined to be a pair of predictable processes (Y,Z) such that almost surely we have $\int_0^T |Z_s|^2 ds < \infty$, $\int_0^T |f(s,Y_s,Z_s)| ds < \infty$, and for all $t \in [0,T]$

$$Y_t = \xi - \int_t^T Z_s dW_s + \int_t^T f(s,Y_s,Z_s) ds.$$

The solution processes (Y,Z) are often shown to satisfy some integrability properties. To this end one usually verifies whether they belong to the following function spaces. Let $p \geq 1$. We denote by $\mathbb{H}^p(\mathbb{R}^3)$ the set of all $\mathbb{R}^3$-valued predictable processes ζ such that $E \int_0^T |\zeta_t|^p dt < \infty$, and by $\mathbb{S}^p(\mathbb{R})$ the set of all $\mathbb{R}$-valued predictable processes δ satisfying $E(\sup_{s \in [0,T]} |\delta_s|^p) < \infty$.

3.3.1. Deriving the FS decomposition with BSDEs

As stated in Section 3.2 we use the framework of local risk minimization. Accordingly, note first that X satisfies (SC), that is X is a special semimartingale with canonical decomposition given by $M = \int \sigma_X X dW$, $\lambda = \frac{\mu_X - r}{\sigma_X^2 X}$ and hence $K_t = \left(\frac{\mu_X - r}{\sigma_X}\right)^2 t$, for $t \in [0,T]$. In order to find the FS decomposition we consider a BSDE with terminal condition $h(I_T)$, and driver f to be specified later,

$$Y_t = h(I_T) - \int_t^T Z_s dW_s + \int_t^T f(s, Y_s, Z_s) ds. \tag{3.11}$$

Assume that f can be chosen such that

$$\int_0^t \xi dX_s = \int_0^t Z_s^1 dW_s^1 - \int_0^t f(s, Y_s, Z_s) ds, \tag{3.12}$$

for all $t \in [0,T]$. Also, by using (3.3), we have

$$\int_0^t \xi dX_s = \int_0^t \xi \sigma_X X_s dW_s^1 + \int_0^t \xi X_s (\mu_X - r) ds. \tag{3.13}$$

Uniqueness of semimartingale decompositions yields that the martingale parts of (3.12) and (3.13) coincide, and therefore it must hold $\xi = \frac{Z^1}{\sigma_X X}$, $P \otimes \lambda$-almost surely. Moreover, the driver f has to satisfy

$$f(s, y, z) = -z^1 \frac{\mu_X - r}{\sigma_X}. \tag{3.14}$$

Indeed, one can show that the solution of the BSDE with generator (3.14) provides the FS decomposition. We summarize this in the next result, which is in fact a special case of El Karoui et al. [1997b, Proposition 1.1].

Lemma 3.9. *The FS decomposition of $h(I_T)$ is given by*

$$h(I_T) = Y_0 + \int_0^T \frac{Z_s^1}{\sigma_X X_s} dX_s + \int_0^T Z_s^2 dW_s^2 + \int_0^T Z_s^3 dW_s^3,$$

where (Y, Z) is the solution of the BSDE (3.11) *with generator f defined as in* (3.14)*.*

In order to obtain a characterization of the solution of

$$Y_t = h(I_T) - \int_t^T Z_s dW_s - \int_t^T Z_s^1 \frac{\mu_X - r}{\sigma_X} ds,$$

we consider the conditional forward SDE given by (3.4) and (3.5) and the associated conditional BSDE

$$Y_s^{t,y,v} = h\left(I_T^{t,y,v}\right) - \int_s^T Z_u^{t,y,v} dW_u - \int_s^T Z_u^{1,t,y,v} \frac{\mu_X - r}{\sigma_X} du, \tag{3.15}$$

for $s \in [t,T]$. Since the BSDE (3.15) is linear we know by standard results, see for example El Karoui et al. [1997b], that

$$Y_s^{t,y,v} = E\left[h\left(I_T^{t,y,v}\right)\Gamma_T^s | \mathcal{F}_s\right] = E\left[h\left(I_T^{t,y,v}\right)\frac{\Gamma_T^0}{\Gamma_s^0} \,\Big|\, \mathcal{F}_s\right],$$

where $\Gamma_s^t = \exp\left(-\frac{\mu_X - r}{\sigma_X}(W_s^1 - W_t^1) - \frac{(\mu_X - r)^2}{2\sigma_X^2}(s-t)\right)$ is the solution of

$$d\Gamma_s^t = \Gamma_s^t\left[-\frac{\mu_X - r}{\sigma_X} dW_s^1\right], \qquad \Gamma_t^t = 1.$$

Let $\tilde{P}$ be the probability measure with density $\frac{d\tilde{P}}{dP} = \Gamma_T^0$, and denote with $\psi : [0,T] \times \mathbb{R}^+ \times (-1,1) \to \mathbb{R}$ the function defined by

$$\psi(t,y,v) = \tilde{E}\left[h\left(I_T^{t,y,v}\right)\right]. \tag{3.16}$$

That the function ψ is well defined and that is has first derivatives with respect to y and v follows from Sections 3.3.2 and 3.3.3. The value process of the solution of the BSDE (3.15) satisfies $Y_s^{t,y,v} = \psi(s, I_s^{t,y,v}, \rho_s^{t,v})$. Our main goal is to derive the explicit hedge formula (3.9). With the help of Lemma 3.17 we get the following representation for $Z_s^{t,y,v}$.

$$Z_s^{t,y,v} = \sigma(I_s^{t,y,v}, \rho_s^{t,v})^* \begin{pmatrix} \partial_y \psi(s, I_s^{t,y,v}, \rho_s^{t,v}) \\ \partial_v \psi(s, I_s^{t,y,v}, \rho_s^{t,v}) \end{pmatrix},$$

where the volatility matrix $\sigma(y,v)$ is given by

$$\sigma(y,v) = \begin{pmatrix} y\sigma_I v & y\sigma_I\sqrt{1-v^2} & 0 \\ g(v)\gamma & g(v)\delta & g(v)\sqrt{1-\gamma^2-\delta^2} \end{pmatrix}, \qquad y \in \mathbb{R}^+, v \in (-1,1).$$

Hence, we have

$$Z_s^{1,t,y,v} = I_s^{t,y,v}\sigma_I \rho_s^{t,v} \partial_y \psi(s, I_s^{t,y,v}, \rho_s^{t,v}) + g(\rho_s^{t,v})\gamma \partial_v \psi(s, I_s^{t,y,v}, \rho_s^{t,v}), \tag{3.17}$$

that is the hedge formula is given by

$$\xi(t,y,x,v) = \frac{y\sigma_I v \partial_y \psi(t,y,v) + g(v)\gamma\partial_v \psi(t,y,v)}{\sigma_X x}. \tag{3.18}$$

Thus by plugging in the explicit representations of $\partial_y \psi(s,y,v)$ and $\partial_v \psi(s,y,v)$, given in Section 3.3.3, we obtain (3.9), and hence have proven Theorem 3.4.

Remark 3.10. Note, that the approach we take by characterizing the Föllmer-Schweizer decomposition via the solution of a linear BSDE is the same as in El Karoui et al. [1997b, Example 1.3]. In our model, however, the inverse of the volatility matrix of the asset processes X and I is unbounded and hence does not fall within the specifications of El Karoui et al. [1997b, Hypothesis 1.1]. Moreover, the coefficients of the volatility matrix of the forward processes I and ρ associated with the BSDE do not satisfy the prerequisites of El Karoui et al. [1997b, Proposition 5.9], that is we do not have uniformly bounded derivatives. In order to recover our hedge formula in spite of these extensions we apply the results of Sections 3.3.2 to 3.4. ⧫

3.3.2. Differentiability with respect to the initial conditions

In this section we want to make some remarks on the system of SDEs, given by Equations (3.4) to (3.7), concerning existence, uniqueness, continuity and differentiablity with respect to the initial values y and v. We also observe the following.

Lemma 3.11. *Suppose that (H1) holds. Then the SDE for $\bar{I}^{t,y,v}$ in* (3.6) *has a unique solution which is given by*

$$\bar{I}_s^{t,y,v} = y\mathcal{E}\left(G^{t,v}\right)_s \int_t^s \mathcal{E}\left(G^{t,v}\right)_u^{-1} dH_u^{t,1,v}, \tag{3.19}$$

for $s \in [t,T]$, where

$$H_s^{t,y,v} = \int_t^s I_u^{t,y,v}\sigma_I \left(\bar{\rho}_u^{t,v} dW_u^1 - \frac{\rho_u^{t,v}}{\sqrt{1-\left(\rho_u^{t,v}\right)^2}}\bar{\rho}_u^{t,v} dW_u^2\right),$$

for $s \in [t,T]$, and

$$G_s^{t,v} = \int_t^s (\mu_I - r)\, du + \int_t^s \sigma_I \left(\rho_u^{t,v} dW_u^1 + \sqrt{1-\left(\rho_u^{t,v}\right)^2} dW_u^2\right),$$

for $s \in [t,T]$.

Proof. Due to (H1) we can define the semimartingales $(H_s^{t,y,v})_{t\le s\le T}$ and $(G_s^{t,v})_{t\le s\le T}$ as above. The dynamics in (3.6) immediately imply that $\bar{I}_s^{t,y,v}$ is the solution of the linear stochastic equation

$$\bar{I}_s^{t,y,v} = H_s^{t,y,v} + \int_t^s \bar{I}_u^{t,y,v} dG_u^{t,v},$$

the solution of which is given by

$$\bar{I}_s^{t,y,v} = \mathcal{E}\left(G^{t,v}\right)_s \left(H_t^{t,y,v} + \int\limits_t^s \mathcal{E}\left(G^{t,v}\right)_u^{-1} \left(dH_u^{t,y,v} - d\left\langle H^{t,y,v}, G^{t,v}\right\rangle_u\right) \right),$$

see for example Revuz and Yor [1999, Chapter IX]. Notice that

$$d\left\langle H^{t,y,v}, G^{t,v}\right\rangle_u = I_u^{t,y,v}\sigma_I^2 \bar{\rho}_u^{t,v}\rho_u^{t,v}du - I_u^{t,y,v}\sigma_I^2\sqrt{1-(\rho_u^{t,v})^2}\frac{\rho_u^{t,v}}{\sqrt{1-(\rho_u^{t,v})^2}}\bar{\rho}_u^{t,v}du = 0.$$

Since $H_t^{t,y,v} = 0$ and $H^{t,y,v} = yH^{t,1,v}$ we obtain (3.19). □

Remark 3.12. The process $I^{t,y,v}$ is given by

$$I_s^{t,y,v} = y\exp\left[\int\limits_t^s \sigma_I\left(\rho_u^{t,v}dW_u^1 + \sqrt{1-\left(\rho_u^{t,v}\right)^2}dW_u^2\right) + \int\limits_t^s\left(-\frac{1}{2}\sigma_I^2 + \mu_I - r\right)du\right],$$

for $s \in [t,T]$. Moreover suppose that $\int_t^s g'(\rho_u^{t,v})d\hat{W}_u$ is well defined, then the SDE in (3.7) has a unique solution $\bar{\rho}^{t,v}$ given by

$$\bar{\rho}_s^{t,v} = \mathcal{E}\left(\int\limits_t^s g'\left(\rho_u^{t,v}\right)d\hat{W}_u + \int\limits_t^s a'\left(\rho_u^{t,v}\right)du\right),$$

for $s \in [t,T]$. ♦

Before we end this section we want to give an auxiliary result which will be used in the sequel.

Lemma 3.13. *Consider two predictable processes c^v and d^v, depending on a parameter $v \in (-1,1)$. Suppose that there exists a continuous function $D : (-1,1) \to \mathbb{R}_+$ such that $c^v + |d^v| \le D(v)$, for all $v \in (-1,1)$. Then the process*

$$b_s^{t,v} = \mathcal{E}\left(\int\limits_t^s d_u^v d\hat{W}_u + \int\limits_t^s c_u^v du\right),$$

is defined, and for all $p \ge 1$ and $v_0 \in (-1,1)$ there exists an open neighbourhood $U \subset (-1,1)$ of v_0, such that

$$\sup_{v\in U} E\left(\sup_{t\le u\le T}|b_u^{t,v}|^p\right) < \infty.$$

Proof. Let $p \ge 1$. The Burkholder-Davis-Gundy Inequality implies that for a constant

C_p, depending only on p, we have,

$$
\begin{aligned}
E\left(\sup_{t\le u\le T}|b_u^{t,v}|^p\right) &\le e^{p(T-t)D(v)}E\left(\sup_{t\le u\le T}\left|\mathcal{E}\left(\int_t^u d_w^v d\hat{W}_w\right)\right|^p\right)\\
&\le C_p e^{p(T-t)D(v)}E\left|\int_t^T \left|(d_u^v)^2\,\mathcal{E}\left(\int_t^u d_w^v d\hat{W}_w\right)\right|^2 du\right|^{p/2}\\
&\le C_p e^{p(T-t)D(v)}D^p(v)E\left|\int_t^T \left|\mathcal{E}\left(\int_t^u d_w^v d\hat{W}_w\right)\right|^2 du\right|^{p/2}.
\end{aligned}
$$

In the rest of the proof we have to distinguish between $p\ge 2$ and $p\in[1,2)$. We first consider $p\ge 2$. By Jensen's inequality and Fubini's theorem we get

$$
\begin{aligned}
E\left|\int_t^T \left|\mathcal{E}\left(\int_t^u d_w^v d\hat{W}_w\right)\right|^2 du\right|^{p/2} &\le E\int_t^T \left|\mathcal{E}\left(\int_t^u d_w^v d\hat{W}_w\right)\right|^p du\\
&= \int_t^T E\left|\mathcal{E}\left(\int_t^u d_w^v d\hat{W}_w\right)\right|^p du. \qquad (3.20)
\end{aligned}
$$

Notice that $|\mathcal{E}(\int_t^u d_w^v d\hat{W}_w)|^p = \exp(\int_t^u p d_w^v d\hat{W}_w + \int_t^u p^2 (d_w^v)^2 dw - \int_t^u p^2(d_w^v)^2 dw - \frac{p}{2}\int_t^u (d_w^v)^2 dw)$, and thus Hölder's inequality implies that the left hand side of Inequality (3.20) can be further estimated against

$$
\begin{aligned}
&E\left|\int_t^T \left|\mathcal{E}\left(\int_t^u d_w^v d\hat{W}_w\right)\right|^2 du\right|^{p/2}\\
&\quad\le \int_t^T \left[E\left(\exp\left(\int_t^u 2p d_w^v d\hat{W}_w - \frac{1}{2}\int_t^u 4p^2\left(d_w^v\right)^2 dw\right)\right)\right]^{1/2}\\
&\qquad\qquad \times\left[E\left(\exp\left(\int_t^u 2p^2\left(d_w^v\right)^2 dw - p\int_t^u \left(d_w^v\right)^2 dw\right)\right)\right]^{1/2} du
\end{aligned}
$$

Consequently,

$$E\left|\int_t^T\left|\mathcal{E}\left(\int_t^u d_w^v d\hat{W}_w\right)\right|^2 du\right|^{p/2} \leq \int_t^T e^{p^2(T-t)D(v)}\left[E\left(\mathcal{E}\left(\int_t^u 2pd_w^v d\hat{W}_w\right)\right)\right]^{1/2} du.$$
$$= (T-t)e^{p^2(T-t)D(v)},$$

which yields

$$E\left(\sup_{t\leq u\leq T}\left|b_u^{t,v}\right|^p\right) \leq C_p D^p(v)(T-t)e^{(p+p^2)(T-t)D(v)},$$

from which we deduce the result for $p \geq 2$. For $1 \leq p < 2$ we use Jensen's inequality to obtain

$$E\left(\sup_{t\leq u\leq T}\left|b_u^{t,v}\right|^p\right) \leq CE\left|\int_t^T\left|\mathcal{E}\left(\int_t^u d_w^{t,v} d\hat{W}_w\right)\right|^2 du\right|^{p/2},$$

and continue with the same arguments as for $p > 2$. Hence the result. □

Remark 3.14. Since $I^{t,y,v}$ is lognormally distributed, independent of v, we have

$$\sup_{v\in(-1,1)} E\left(\left|I_s^{t,y,v}\right|^p\right) < \infty,$$

for all $p \geq 1$. Let K be a compact subset of $(-1,1)$ and suppose $\sup_{v\in K} g'(\rho_s^{t,v})$ is bounded and $\sup_{v\in K} a'(\rho_s^{t,v})$ is bounded above, uniformly for all $t \leq s \leq T$. Then we have $\sup_{v\in K} E(\int_t^T |\bar{\rho}_s^{t,v}|^p ds) < \infty$, for all $p \geq 1$, by Lemma 3.13. ♦

3.3.3. Differentiability of ψ

In order to derive the hedge formula (3.18) we need to ensure that ψ defined in (3.16) is continuously differentiable with respect to v and y. We only consider the differentiablity in v, since for y it is comparatively simpler. Since we want to use uniform integrability this is where (H1) and (H2) come into play.

Lemma 3.15. *Suppose (H1) and (H2) hold. Then, for all $v_0 \in (-1,1)$, there exists an open neighbourhood $U \subset (-1,1)$, such that*

$$\sup_{v\in U} E\left(\left|\bar{I}_T^{t,y,v}\right|^{p'}\right) < C, \tag{3.21}$$

for all $p' \in [1,p)$, with p of Assumption (H2). Moreover C is a constant that depends only on p from Condition (H2), the model's parameters and U.

Proof. Let $p' \geq 1$, such that $p > p'$ with p of Assumption (H2). Let $G^{t,v}$ and $H^{t,y,v}$ be defined as in Lemma 3.11. Notice that $G^{t,v}$ is lognormally distributed. Since the distribution does not depend on the correlation, there exists a constant $C \in \mathbb{R}_+$ such that we have $E(|\mathcal{E}(G^{t,v})|^{2p'}) \leq C$, for all $v \in (-1,1)$. The Cauchy-Schwarz Inequality, the Burkholder-Davis-Gundy Inequality and Jensen's inequality imply

$$\begin{aligned}
E\left(\left|\bar{I}_T^{t,y,v}\right|^{p'}\right) &= E\left(\left|\mathcal{E}\left(G^{t,v}\right)_T \int_t^T \mathcal{E}\left(G^{t,v}\right)_u^{-1} dH_u^{t,y,v}\right|^{p'}\right) \\
&\leq \sqrt{C}\left(E\left(\sup_{t\leq s\leq T}\left|\int_t^s \mathcal{E}\left(G^{t,v}\right)_u^{-1} dH_u^{t,y,v}\right|^{2p'}\right)\right)^{\frac{1}{2}} \\
&\leq \sqrt{C}C_{p'}\left(E\left(\left|\int_t^T \mathcal{E}\left(G^{t,v}\right)_u^{-2}\left(I_u^{t,y,v}\right)^2 \sigma_I^2\left[\frac{\left(\bar{\rho}_u^{t,v}\right)^2}{1-\left(\rho_u^{t,v}\right)^2}\right] du\right|^{p'}\right)\right)^{\frac{1}{2}} \\
&\leq \sqrt{C}C_{p'}\left(E\left(\left|\int_t^T \mathcal{E}\left(G^{t,v}\right)_u^{-2p'}\left(I_u^{t,y,v}\right)^{2p'} \sigma_I^{2p'}\left[\frac{\left(\bar{\rho}_u^{t,v}\right)^2}{1-\left(\rho_u^{t,v}\right)^2}\right]^{p'} du\right|\right)\right)^{\frac{1}{2}}.
\end{aligned}$$

Now choose $\hat{p} > 1$, such that $\hat{p}p' < p$. An application of the Hölder Inequality yields, with $\hat{q} = \frac{\hat{p}}{\hat{p}-1}$,

$$\begin{aligned}
\left(E\left(\left|\bar{I}_T^{t,y,v}\right|^{p'}\right)\right)^2 \leq CC_{p'}^2 &\left(E\left(\left|\int_t^T \mathcal{E}\left(G^{t,v}\right)_u^{-2p'\hat{q}}\left(I_u^{t,y,v}\right)^{2p'\hat{q}} du\right|\right)\right)^{1/\hat{q}} \\
&\times\left(E\left(\left|\int_t^T \left[\frac{\left(\bar{\rho}_u^{t,v}\right)^2}{1-\left(\rho_u^{t,v}\right)^2}\right]^{p'\hat{p}} du\right|\right)\right)^{1/\hat{p}}.
\end{aligned}$$

For any $U \subset (-1,1)$, the supremum $\sup_{v\in U} E(|\int_t^T \mathcal{E}(G^{t,v})_u^{-2p'\hat{q}}(I_u^{t,y,v})^{2p'\hat{q}}du|)$ is finite due to the lognormal distribution of $I^{t,y,v}$ and the normal distribution of $G^{t,v}$, the distributions of which do not depend on the correlation process ρ. Therefore, with U from Assumption (H2), we get $\sup_{v\in U} E(|\bar{I}_T^{t,y,v}|^{p'}) < C$. □

The following lemma states conditions under which ψ is differentiable with respect to v.

Lemma 3.16. *Let h be Lipschitz such that the weak derivative h' is Lebesgue-almost everywhere continuous. Under (H1) and (H2) $\psi(t,y,v)$ is continuously differentiable with respect to v and the partial derivative $\partial_v\psi(t,y,v)$ is given by*

$$\partial_v\psi(t,y,v) = \tilde{E}\left[h'\left(I_T^{t,y,v}\right)\bar{I}_T^{t,y,v}\right].$$

Proof. Let v_0 be an element of $(-1,1)$, and $p>1$ as in (H2). According to Lemma 3.15 we can choose a $\delta>0$ with $(v_0-\delta, v_0+\delta)\subset(-1,1)$, such that, for all $p'\in[1,p)$, we have $\sup_{v\in(v_0-\delta,v_0+\delta)} E(|\bar{I}_T^{t,y,v}|^{p'})<\infty$. For all $v\in(v_0-\delta,v_0+\delta)$, we will show

1. $\psi(t,y,v)$ is well defined,

2. $h\left(I_T^{t,y,v}\right)$ is absolutely continuous in v,

3. $\tilde{E}\left[h'\left(I_T^{t,y,v}\right)\bar{I}_T^{t,y,v}\right]$ is continuous at $v=v_0$, and

4. $\tilde{E}\left[\int_{-\delta}^{\delta}\left|h'\left(I_T^{t,y,v_0+u}\right)\bar{I}_T^{t,y,v_0+u}\right|du\right]<\infty$.

By standard arguments these four statements imply the result, see for instance Durrett [1996].

The properties of h imply that there exists a constant $C\in\mathbb{R}_+$ such that $|h(x)|\le C(1+|x|)$, and hence with Remark 3.14 we have, with $q=\frac{p}{p-1}$,

$$\begin{aligned}\tilde{E}\left[\left|h\left(I_T^{t,y,v}\right)\right|\right]&\le E\left(\left|\Gamma_T^0\right|^q\right)^{1/q} E\left(\left|h\left(I_T^{t,y,v}\right)\right|^p\right)^{1/p}\\&\le E\left(\left|\Gamma_T^0\right|^q\right)^{1/q}\ 2C\left(1+E\left|I_T^{t,y,v}\right|^p\right)^{1/p}\\&<\infty,\end{aligned}$$

and therefore ψ is well defined.

Since h is Lipschitz, it is absolutely continuous. Besides, $I_T^{t,y,v}$ is differentiable and continuous in v (see Section 3.2), and consequently, the composition $h(I_T^{t,y,v})$ is absolutely continuous in v. With Hölder's inequality we have, for $p'\in[1,p)$ and $\hat{p}>1$, such that $p>p'\hat{p}>1$,

$$\tilde{E}\left[\left|h'\left(I_T^{t,y,v}\right)\bar{I}_T^{t,y,v}\right|^{p'}\right]\le C\left(E\left[\left|\bar{I}_T^{t,y,v}\right|^{p'\hat{p}}\right]\right)^{1/\hat{p}}.$$

Thus, by Lemma 3.15 we get

$$\sup_{v\in(v_0-\delta,v_0+\delta)}\tilde{E}\left[\left|h'\left(I_T^{t,y,v}\right)\bar{I}_T^{t,y,v}\right|^{p'}\right]\le C<\infty.$$

Hence, the family of random variables $(h'(I_T^{t,y,v})\bar{I}_T^{t,y,v})_{v\in(v_0-\delta,v_0+\delta)}$ is uniformly integrable with respect to $\tilde{P}$. Now, let $(v_n)_{n\in\mathbb{N}}$ be any sequence in $(v_0-\delta,v_0+\delta)$ with limit v_0. Then by continuity of $h'(I_T^{t,y,v})\bar{I}_T^{t,y,v}$ and the properties of uniform integrability we get

$$\begin{aligned}&\lim_{n\to\infty}\left|\tilde{E}\left[h'\left(I_T^{t,y,v_n}\right)\bar{I}_T^{t,y,v_n}\right]-\tilde{E}\left[h'\left(I_T^{t,y,v_0}\right)\bar{I}_T^{t,y,v_0}\right]\right|\\&\qquad\le\lim_{n\to\infty}\tilde{E}\left[\left|h'\left(I_T^{t,y,v_n}\right)\bar{I}_T^{t,y,v_n}-h'\left(I_T^{t,y,v_0}\right)\bar{I}_T^{t,y,v_0}\right|\right]=0,\end{aligned}$$

that is the continuity of $\tilde{E}[h'(I_T^{t,y,v})\bar{I}_T^{t,y,v}]$ at $v = v_0$. We use boundedness of h' and Fubini's Theorem to get

$$\tilde{E}\left[\int_{-\delta}^{\delta} \left|h'\left(I_T^{t,y,v_0+u}\right)\bar{I}_T^{t,y,v_0+u}\right| du\right] \le C\int_{-\delta}^{\delta} E\left|\bar{I}_T^{t,y,v_0+u}\right| du \le C \sup_{v\in(v_0-\delta,v_0+\delta)} E\left|\bar{I}_T^{t,y,v}\right|,$$

which is finite by Lemma 3.15. Since we verified 1.-4. the proof of Lemma 3.16 is complete. □

3.3.4. The hedge as variational derivative

The control process $Z^{t,y,v}$ of the linear BSDE (3.15) has a representation in terms of the gradient of ψ and the matrix-valued function defined by

$$\sigma(y,v) = \begin{pmatrix} y\sigma_I v & y\sigma_I\sqrt{1-v^2} & 0 \\ g(v)\gamma & g(v)\delta & g(v)\sqrt{1-\gamma^2-\delta^2} \end{pmatrix}, \qquad y\in\mathbb{R}^+, v\in(-1,1).$$

Lemma 3.17. *Assume that (H1) and (H2) hold, that a and g are continuously differentiable and let h be Lipschitz such that the weak derivative h' is Lebesgue-almost everywhere continuous. Then, for $s\in[t,T]$,*

$$Z_s^{t,y,v} = \sigma(I_s^{t,y,v},\rho_s^{t,v})^* \begin{pmatrix} \partial_y\psi(s,I_s^{t,y,v},\rho_s^{t,v}) \\ \partial_v\psi(s,I_s^{t,y,v},\rho_s^{t,v}) \end{pmatrix}. \tag{3.22}$$

Proof. By Lemma 3.16, $\psi(s,y,v)$ is continuously differentiable in y and v. As is shown in Imkeller et al. [2012, Theorem 5.2 and Remark 5.3.i] this is sufficient for Equation (3.22) to hold. □

Note that in Imkeller et al. [2012] the authors establish representations as in (3.22) by using only elementary methods. However, up to now the standard method of deriving these relationships was to interpret $Z^{t,y,v}$ as the Malliavin derivative, or more precisely the Malliavin trace, of $Y^{t,y,v}$. Compared to the approach given in Imkeller et al. [2012] this has the disadvantage that additional regularity assumptions which originate in the usage of the Malliavin calculus are needed. Nevertheless we want to outline how Malliavin calculus can be used to derive (3.22), thus giving a proof of (3.22) in this thesis (though not in full generality). Since this approach entails variational derivatives of the forward processes I and ρ, see Equation (3.23), we need the additional assumption that the coefficients a and g of the dynamics of ρ have bounded derivatives.

Proof (under the additional assumptions that a and g have bounded derivatives). Let I^n be the solution of the SDE

$$dI_t^n = I_t^n(\mu_I - r)dt + \sigma_I I_t^n\left(\left(1-\frac{1}{n}\right)\rho_t dW_t^1 + \sqrt{1-\left(1-\frac{1}{n}\right)^2\rho_t^2}\,dW_t^2\right).$$

It is straigtforward to show that I_T^n converges to I_T in L^2. By taking a subsequence, we may assume that I_T^n converges to I_T almost surely.

Next we approximate the payoff function h by a sequence of everywhere differentiable and globally Lipschitz continuous functions. More precisely, let $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$, the density of a standard normal distibution and let $\varphi^n(x) = n\varphi(nx)$, $x \in \mathbb{R}$, for all $n \geq 1$. We define h^n as the convolution with φ, that is $h^n = h * \varphi^n$. Observe that h^n is Lipschitz continuous with respect to the same Lipschitz constant as h. Note that Lipschitz continuity of h implies uniform convergence of h^n to h, hence $h^n(I_T^n)$ converges P-almost surely to $h(I_T)$. Moreover, h^n is differentiable.

As before, we denote by $I^{n,t,y,v}$ the process I^n conditioned on $I_t^n = y$ and $\rho_t = v$. We further define $\psi^n(t, y, v) = \tilde{E}[h^n(I_T^{n,t,y,v})]$, for all $n \geq 1$, where $\tilde{E}$ denotes the expectation with respect to the measure $\tilde{P}$ defined in Section 3.2. Note that by the same methods as in Section 3.3.3 it can be shown that the ψ^n are differentiable; indeed, due to the factor $(1 - \frac{1}{n})$ the integrability condition in (3.21) is trivial. Moreover, its derivatives are bounded, that is the ψ^n are Lipschitz continuous.

We proceed by showing that ψ^n converges pointwise to ψ. Indeed, with $L \in \mathbb{R}_+$ being the Lipschitz constant of h, we have

$$
\begin{aligned}
\lim_n |\psi^n(t, y, v) - \psi(t, y, v)| &\leq \lim_n \sqrt{\tilde{E}\left|h^n\left(I_T^{n,t,y,v}\right) - h\left(I_T^{t,y,v}\right)\right|^2} \\
&\leq \lim_n \sqrt{4\left(\tilde{E}\left|h^n\left(I_T^{n,t,y,v}\right) - h\left(I_T^{n,t,y,v}\right)\right|^2 + \tilde{E}\left|h\left(I_T^{n,t,y,v}\right) - h\left(I_T^{t,y,v}\right)\right|^2\right)} \\
&\leq \lim_n \sqrt{4\left(\|h^n - h\|_\infty + L^2 \tilde{E}\left|I_T^n - I_T\right|^2\right)} = 0.
\end{aligned}
$$

Let $(Y^{n,t,y,v}, Z^{n,t,y,v})$ be the solution of the BSDE

$$
Y_s^{n,t,y,v} = h^n\left(I_T^{n,t,y,v}\right) - \int_s^T Z_u^{n,t,y,v} dW_u - \int_s^T Z_u^{n,1,t,y,v} \frac{\mu_X - r}{\sigma_X} du,
$$

for $s \in [t, T]$. Since $h^n(I_T^{n,t,y,v})$ converges to $h(I_T^{t,y,v})$ in $L^2(P)$, standard a priori estimates for Lipschitz BSDEs, or simply the Ito isometry under the measure $\tilde{P}$, imply that (Y^n, Z^n) converges to (Y, Z) in $\mathbb{S}(\mathbb{R}) \otimes \mathbb{H}_T^2(\mathbb{R}^3)$.

Notice that, due to the Markov property, we have $Y_s^{n,t,y,v} = \tilde{E}[h^n(I_T^{n,t,y,v})|\mathcal{F}_s] = \psi^n(s, I_s^{n,t,y,v}, \rho_s^{t,v})$. Since the approximations ψ^n are Lipschitz continuous, we may apply the chain rule, which yields

$$
D_u Y_s^{n,t,y,v} = \partial_y \psi^n\left(s, I_s^{n,t,y,v}, \rho_s^{t,v}\right) D_u I_s^{n,t,y,v} + \partial_v \psi^n\left(s, I_s^{n,t,y,v}, \rho_s^{t,v}\right) D_u \rho_s^{t,v}, \quad (3.23)
$$

for $u \in [t, T]$, where D_u denotes the Malliavin derivative of $Y^{n,t,y,v}$, $I^{n,t,y,v}$ and $\rho^{t,v}$ respectively. $D_u I^{n,t,y,v}$ and $D_u \rho^{t,v}$ are solutions of linear SDEs, see Nualart [2006, Theorem 2.2.1]. In particular this guarantees right continuity of $D_u Y_s^{n,t,y,v}$ in s.

By the Clark-Ocone formula, the process $Z^{n,t,y,v}$ is the predictable projection of

$h^n(I_T^{n,t,y,v})$ under the measure $\tilde{P}$. More precisely, for all $s \in [t,T]$, we have $Y_s^{n,t,y,v} = \int_t^s \tilde{E}[D_u Y_s^{n,t,y,v}|\mathcal{F}_u] d\tilde{W}_u$, where $\tilde{W}_t = (W_t^1 + \frac{\mu_X - r}{\sigma_X} t, W_t^2, W_t^3)$ and $\tilde{E}[\cdot|\mathcal{F}_u]$ stands for the predictable projection operator with respect to $\tilde{P}$. Due to the right continuity of $D_u Y_s^{n,t,y,v}$ in s, we may interchange the Malliavin and the predictable projection operator, which yields

$$\begin{aligned} Z_u^{n,t,y,v} = \tilde{E}\left[D_u Y_s^{n,t,y,v}|\mathcal{F}_u\right] &= \lim_{s\downarrow u} D_u Y_s^{n,t,y,v} \\ &= \sigma^n\left(u, I_u^{n,t,y,v}, \rho_u^{t,v}\right)^* \begin{pmatrix} \partial_y \psi^n(u, I_u^{n,t,y,v}, \rho_u^{t,v}) \\ \partial_v \psi^n(u, I_u^{n,t,y,v}, \rho_u^{t,v}) \end{pmatrix}, \end{aligned} \tag{3.24}$$

where

$$\sigma^n(y,v) = \begin{pmatrix} y\sigma_I(1-\frac{1}{n})v & y\sigma_I\sqrt{1-(1-\frac{1}{n})^2 v^2} & 0 \\ g(v)\gamma & g(v)\delta & g(v)\sqrt{1-\gamma^2-\delta^2} \end{pmatrix}.$$

We next show that the partial derivatives $\partial_y\psi^n$ and $\partial_v\psi^n$ converge pointwise to $\partial_y\psi$ and $\partial_v\psi$, respectively. To this end denote again the derivatives of $I^{n,t,y,v}$ with respect to v by $\bar{I}^{n,t,y,v}$. Lemma 3.15 yields that $\sup_n E|\bar{I}_T^n|^p < \infty$, which further implies that the sequence $(\bar{I}_T^n)$ is uniformly integrable. Moreover,

$$\begin{aligned} |\partial_v\psi^n - \partial_v\psi| &\le \tilde{E}\left|(h^n)'\left(I_T^{n,t,y,v}\right)\bar{I}_T^{n,t,y,v} - h'\left(I_T^{t,y,v}\right)\bar{I}_T^{t,y,v}\right| \\ &\le \tilde{E}\left|(h^n)'\left(I_T^{n,t,y,v}\right)\right|\left|\bar{I}_T^{n,t,y,v} - \bar{I}_T^{t,y,v}\right| + \tilde{E}\left|\bar{I}_T^{t,y,v}\right|\left|(h^n)'\left(I_T^{n,t,y,v}\right) - h'\left(I_T^{t,y,v}\right)\right|. \end{aligned} \tag{3.25}$$

We show separately that both summands in (3.25) converge to 0 as $n \to \infty$. Since the approximating functions h^n have one common Lipschitz constant $L \in \mathbb{R}_+$, the derivatives satisfy $|(h^n)'| \le L$, for all $n \ge 1$. Consequently, $\lim_n \tilde{E}|(h^n)'(I_T^{n,t,y,v})||\bar{I}_T^{n,t,y,v} - \bar{I}_T^{t,y,v}| \le L \lim_n \tilde{E}|\bar{I}_T^{n,t,y,v} - \bar{I}_T^{t,y,v}|$. Next, let $\tau_k = T \wedge \inf\{t \ge 0 : |\rho_t| = (1-\frac{1}{k})\}$, for all $k \ge 1$. Then the stopped processes $\bar{I}^n_{\cdot\wedge\tau_k}$ converge to $\bar{I}_{\cdot\wedge\tau_k}$ in L^2 as $n \to \infty$, see for example Protter [2004, Chapter V.4]. Therefore, by dominated convergence, for every $k \ge 1$, we have

$$\begin{aligned} &\lim_n \tilde{E}\left|\bar{I}_T^{n,t,y,v} - \bar{I}_T^{t,y,v}\right| \\ &\quad \le \lim_n \left(\tilde{E}\left|1_{\{\tau_k<T\}}\left(\bar{I}_T^{n,t,y,v} - \bar{I}_T^{t,y,v}\right)\right| + \tilde{E}\left|1_{\{\tau_k\ge T\}}\left(\bar{I}_T^{n,t,y,v} - \bar{I}_T^{t,y,v}\right)\right|\right) \\ &\quad\quad \le \tilde{E}\left(1_{\{\tau_k<T\}}\left(\left|\bar{I}_T^{n,t,y,v}\right| + \tilde{E}\left|\bar{I}_T^{t,y,v}\right|\right)\right). \end{aligned}$$

Recall that $(\bar{I}_T^n)$ is uniformly integrable, and that $\lim_k \tilde{P}(\tau_k = T) = 1$. Hence, by letting $k \to \infty$ we get that $\lim_n \tilde{E}|\bar{I}_T^{n,t,y,v} - \bar{I}_T^{t,y,v}| = 0$, and hence the first summand in (3.25) converges to 0.

In order to show that the second summand in (3.25) vanishes, we first show that, for $x_n \to x$, $(h^n)'(x_n) \to h'(x)$. If x is a point of continuity of h', then we have the following estimate

$$
\begin{aligned}
|(h^n)'(x_n) - h'(x)| &= \left| \int \varphi_n(x_n - y)h'(y)dy - h'(x) \right| \\
&= \left| \int \varphi_n(x_n - x + x - y)h'(y)dy - h'(x) \right| \\
&= \left| \int \varphi_n(x - z)h'(z + x_n - x)dz - h'(x) \right| \qquad (z := y - x_n + x) \\
&\le \left| \int \varphi_n(x - z)h'(z)dz - h'(x) \right| \\
&\quad + \left| \int \varphi_n(x - z)[h'(z + x_n - x) - h'(z)]dz \right|.
\end{aligned}
$$

Applying the transformation $y := n(x - z)$ in each term on the right hand side of the inequality, together with dominated convergence and the continuity of h' in x, yields that $\lim_n h'(x_n) = h'(x)$.

Since I_T has a density, $(h^n)'(I_T^{n,t,y,v})$ converges to $h'(I_T^{t,y,v})$ almost everywhere. Consequently, by dominated convergence, we obtain

$$
\lim_n \tilde{E}|\bar{I}_T^{t,y,v}||(h^n)'(I_T^{n,t,y,v}) - h'(I_T^{t,y,v})| = 0.
$$

Thus we have shown that $\lim_n \partial_v \psi^n(t, y, v) = \partial_v \psi(t, y, v)$, for all $t \in [0, T]$, $y \in \mathbb{R}$ and $v \in (-1, 1)$.

Notice that $I_s^{n,t,y,v} = yI_s^{n,t,1,v}$ and $\bar{I}_s^{n,t,y,v} = y\bar{I}_s^{n,t,1,v}$, which implies that, for all $t \ge 0$ and $v \in (-1, 1)$, the sequence $(\partial_v \psi^n(t, \cdot, v))$ converges to $\partial_v \psi(t, \cdot, v)$ uniformly in y on all compact sets of $\mathbb{R}$. Similarly, one can show locally uniform convergence in y of the partial derivatives $\partial_y \psi^n$ to $\partial_y \psi$.

This finally yields that $(\partial_v \psi^n(s, I_s^{n,t,y,v}, \rho_s^{t,v}))$ converges to $\partial_v \psi(s, I_s^{t,y,v}, \rho_s^{t,v})$, and $(\partial_y \psi^n(s, I_s^{n,t,y,v}, \rho_s^{t,v}))$ to $\partial_y \psi(s, I_s^{t,y,v}, \rho_s^{t,v})$, almost surely. Moreover, by combining this with Equation (3.24), we get that Z^n converges almost surely to $\sigma(s, I_s^{t,y,v}, \rho_s^{t,v})^* \begin{pmatrix} \partial_y \psi(s, I_s^{t,y,v}, \rho_s^{t,v}) \\ \partial_v \psi(s, I_s^{t,y,v}, \rho_s^{t,v}) \end{pmatrix}$. Since Z^n converges also to Z in $\mathbb{H}^2$, we obtain the result. □

3.4. A class of correlation dynamics which fulfill the main assumptions

In this part of the work we characterize a class of dynamics which fulfill Conditions (H1) and (H2). One result has already been mentioned as Theorem 3.8 in Section 3.2 above. We will give its proof below.

Moreover we will give an extension of Theorem 3.8 in Proposition 3.20. We will see in the Section 3.5, that this extension enables us to show that the so-called Jacobi processes also fit into our framework. In contrast to the dynamics given in Theorem 3.8

the diffusion coefficient of a Jacobi processes has unbounded derivatives in -1 and 1.

We first collect some notation and facts on attainability of boundaries for diffusions. The material is taken from Karlin and Taylor [1981]. Suppose we are given a general diffusion

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad l \leq X_0 \leq r,$$

where l (resp. r) denote the left boundary l (resp. the right boundary). In the following we only consider the analysis for the left boundary. We define, for $x \in (l, r)$,

$$\begin{aligned}
s(v) &= \exp\left(-\int_{v_0}^{v} \frac{2\mu(w)}{\sigma^2(w)} dw\right), \qquad v_0 \in (l, x), \\
S(x) &= \int_{x_0}^{x} s(v)dv, \qquad x_0 \in (l, x), \\
S[c, d] &= S(d) - S(c), \qquad (c, d) \in (l, r), \\
S(l, x] &= \lim_{c \to l} S[c, x].
\end{aligned}$$

We already indicate that x_0 and v_0 will be of no relevance in the following. S is called the *scale measure* whereas M is the *speed measure*:

$$\begin{aligned}
m(x) &= \frac{1}{\sigma^2(x)s(x)}, \\
M[c, d] &= \int_{c}^{d} m(x)dx.
\end{aligned}$$

We also need

$$\Sigma(l) = \lim_{c \to l} \int_{c}^{x} M[v, x] dS(v).$$

According to Karlin and Taylor [1981] the boundary l is attracting if $S(l, x] < \infty$ and this criterion applies independently of $x \in (l, r)$. Moreover, the boundary l is said to be

1. attainable if $\Sigma(l) < \infty$,
2. unattainable if $\Sigma(l) = \infty$.

For a proof of the following Lemma see Karlin and Taylor [1981, Chapter 15.6].

Lemma 3.18. *$S(l, x] = \infty$ implies $\Sigma(l) = \infty$.*

With this at hand, we can find sufficient conditions on the coefficients of the correlation dynamics so that (H1) is satisfied. Let again $\rho^{t,v}$ and $\bar{\rho}^{t,v}$ be defined as in (3.5) and (3.7), respectively. To simplify notation, from now on we will suppress the dependence on t and v and only write ρ resp. $\bar{\rho}$.

Lemma 3.19. *Let a and g be continuously differentiable. We assume that g does not have any roots in $(-1,1)$. If*

$$\limsup_{x\uparrow 1} \frac{2a(x)(1-x)}{g^2(x)} < 0 \text{ and } \liminf_{x\downarrow -1} \frac{2a(x)(1+x)}{g^2(x)} > 0, \tag{3.26}$$

then Condition (H1) is satisfied.

Proof. We show that ρ does not reach -1. By (3.26), there exist $\varepsilon > 0$, $\delta > 0$ and $v_0 \in (-1,1)$, such that $\frac{2a(w)}{g^2(w)} \geq \frac{\varepsilon}{1+w}$, for all $-1 < w < v_0 < -1+\delta$. Hence,

$$s(v) = \exp\left(\int_v^{v_0} \frac{2a(w)}{g^2(w)} dw\right) \geq \exp\left(\int_v^{v_0} \frac{\varepsilon}{1+w} dw\right) = C\exp(-\log(1+v)) = C\frac{1}{1+v}.$$

Consequently,

$$S[c,x] \geq C\int_c^x \frac{1}{1+v} dv \longrightarrow \infty,$$

for $c \to -1$, and thus by Lemma 3.18 we obtain that ρ does not reach -1. We treat the boundary 1 similar and hence (H1) holds. $\square$

The next proposition provides conditions under which Condition (H2) is satisfied. We will need two auxiliary processes $\tilde{a}$ and $\tilde{g}$ defined by, for $u \in [0,T]$,

$$\tilde{a}_u = \frac{2\rho_u}{(1-\rho_u^2)} g(\rho_u) + 2g'(\rho_u), \text{ and}$$
$$\tilde{g}_u = \frac{2\rho_u}{(1-\rho_u^2)} a(\rho_u) + 2a'(\rho_u) + \frac{g^2(\rho_u)}{1-\rho_u^2} + \frac{4\rho_u^2 g^2(\rho_u)}{(1-\rho_u^2)^2} + (g'(\rho_u))^2 + \frac{4\rho_u}{1-\rho_u^2} g(\rho_u)g'(\rho_u).$$

Proposition 3.20. *Assume the conditions of Lemma 3.19 are satisfied. Then Assumption (H1) holds, and therefore, $\tilde{a}$ and $\tilde{g}$ are well defined. Suppose $\tilde{a}$ is bounded and $\tilde{g}$ is bounded from above. Then Assumption (H2) is satisfied, and hence, the delta hedge is given as in Theorem 3.4.*

Proof. We start by an application of Ito's formula on the process Φ defined by $\Phi_s = f(\rho_s, \bar{\rho}_s)$, where f is given by $f(x,y) = \frac{y^2}{1-x^2}$. Note that $f_x(x,y) = \frac{2xy^2}{(1-x^2)^2}$, $f_{xx} =$

$\frac{2y^2}{(1-x^2)^2} + \frac{8x^2y^2}{(1-x^2)^3}$, $f_y(x,y) = \frac{2y}{1-x^2}$, $f_{yy} = \frac{2}{1-x^2}$ and $f_{xy} = \frac{4xy}{(1-x^2)^2}$. We have

$$\begin{aligned}
\Phi_s = \Phi_t &+ \int_t^s \frac{2\rho_u\bar{\rho}_u^2}{(1-\rho_u^2)^2}\left[a(\rho_u)du + g(\rho_u)d\hat{W}_u\right] \\
&+ \int_t^s \frac{2\bar{\rho}_u}{1-\rho_u^2}\left[a'(\rho_u)\bar{\rho}_u du + g'(\rho_u)\bar{\rho}_u d\hat{W}_u\right] \\
&+ \frac{1}{2}\int_t^s \left[\frac{2\bar{\rho}_u^2}{(1-\rho_u^2)^2} + \frac{8\rho_u^2\bar{\rho}_u^2}{(1-\rho^2)^3}\right] g^2(\rho_u)du + \frac{1}{2}\int_t^s \frac{2}{1-\rho_u^2}(g'(\rho_u))^2\bar{\rho}_u^2 du \\
&+ \int_t^s \frac{4\rho_u\bar{\rho}_u}{(1-\rho^2)^2} g(\rho_u)g'(\rho_u)\bar{\rho}_u du \\
= \;\; \Phi_t &+ \int_t^s \Phi_u\left[\frac{2\rho_u}{(1-\rho_u^2)}g(\rho_u) + 2g'(\rho_u)\right]d\hat{W}_u \\
&+ \int_t^s \Phi_u\left[\frac{2\rho_u}{(1-\rho_u^2)}a(\rho_u) + 2a'(\rho_u) + \frac{g^2(\rho_u)}{1-\rho_u^2} + \frac{4\rho_u^2 g^2(\rho_u)}{(1-\rho_u^2)^2}\right. \\
&\qquad\left. + (g'(\rho_u))^2 + \frac{4\rho_u}{1-\rho_u^2}g(\rho_u)g'(\rho_u)\right]du.
\end{aligned}$$

Thus, Φ is the solution of a linear stochastic equation and given by

$$\Phi_s = \Phi_t\mathcal{E}\left(\int_t^s \tilde{a}_u d\hat{W}_u + \int_t^s \tilde{g}_u du\right).$$

Hence by our assumptions on $\tilde{a}$ and $\tilde{g}$, and by Lemma 3.13, all moments of $\sup_{t\le u\le T}\Phi_u$ are finite, which further yields (H2). $\square$

We use the two preceding statements to prove Theorem 3.8.

Proof (of Theorem 3.8). Condition (H1) follows from Lemma 3.19. Since 1 and -1 are roots of g we can write $\frac{g(x)}{1-x^2} = \frac{1}{1+x}\frac{g(x)-g(1)}{1-x} = \frac{-1}{1+x}\frac{g(x)-g(1)}{x-1}$ and hence $\frac{g(x)}{1-x^2}$ is bounded for $x \nearrow 1$ by the derivative of g at $x = 1$. Similarly for $x \searrow -1$, and consequently the fraction $\frac{g(x)}{1-x^2}$ is bounded on $[-1,1]$. Moreover, Condition (3.26) implies that there exists an $\varepsilon \in (0,1)$, such that all $x \in (-1,1)$ with $|x| \ge 1-\varepsilon$ satisfy $xa(x) < 0$. Hence $\tilde{a}$ (resp. $\tilde{g}$) is bounded (resp. bounded from above) and therefore we obtain the result by Proposition 3.20. $\square$

Remark 3.21. 1. Note that the conditions on the coefficients a and g of the correlation dynamics in Theorem 3.8 are more restrictive than in Proposition 3.20. This is mainly for ease of exposition in Section 3.2. In Section 3.5.1 an example is given where the coefficient g of the correlation dynamics does not have a bounded derivative.

2. It is possible to prove Theorem 3.8 without considering the auxiliary processes $\tilde{a}$ and $\tilde{g}$ and using Proposition 3.20. In the following we give a rough sketch of a more intuitive proof of Theorem 3.8. That alternative proof consists in showing that all moments of the process $Y_t = \frac{1}{1-\rho_t^2}$ are finite, from which one easily deduces Condition (H2) to be satisfied. From Ito's formula we obtain

$$dY_t = \frac{2\rho_t}{(1-\rho_t^2)} g(\rho_t) Y_t d\hat{W}_t + 2\rho_t a(\rho_t) Y_t^2 dt + \left(1+3\rho_t^2\right) \frac{g^2(\rho_t)}{(1-\rho_t^2)^2} Y_t dt,$$

showing that Y is a linear SDE with an additional drift term growing quadratically in Y. Condition (3.10) implies that there exists an $\varepsilon \in (0,1)$, such that all $x \in (-1,1)$ with $|x| \geq 1-\varepsilon$ satisfy $xa(x) < 0$. Moreover, $\{|\rho_s| \leq 1-\varepsilon\} = \{Y_s \leq \frac{1}{2\varepsilon - \varepsilon^2}\}$, and consequently, the quadratic drift term in the dynamics of Y has a shrinking effect as soon as Y exceeds $C_\varepsilon = \frac{1}{2\varepsilon-\varepsilon^2}$. In other words, Y can be shown to be dominated by the SDE

$$d\check{Y}_t = \frac{2\rho_t}{(1-\rho_t^2)} g(\rho_t) \check{Y}_t d\hat{W}_t + \left(1+3\rho_t^2\right) \frac{g^2(\rho_t)}{(1-\rho_t^2)^2} \check{Y}_t dt + C_\varepsilon dt,$$

that, by standard arguments, can be shown to possess finite moments. ♦

3.5. Examples

The aim of this final section is to give some explicit correlation dynamics which fall within the framework above. We start by modelling correlation processes directly as solutions of various SDEs with values in $[-1,1]$ in Subsection 3.5.1. Another approach is used in Subsection 3.5.2 where we use mappings of an Ornstein-Uhlenbeck process onto the open intervall $(-1,1)$.

3.5.1. Modelling correlation directly

Example 3.22. Of course all processes that are bounded away from -1 and 1 fulfill the Conditions (H1) and (H2). ◊

Example 3.23. For $a(x) = \kappa(\theta - x)$, with $\kappa > 0$, $\theta \in (-1,1)$, and $g(x) = \alpha(1-x^2)$ in the dynamics of ρ, the prerequisites of Theorem 3.8 are fulfilled. ◊

Example 3.24. Let a and g be polynomials. Assume that $g(-1) = g(1) = 0$, and that g does not have any roots in $(-1,1)$. If

$$\lim_{x\uparrow 1} \frac{a(x)}{g^2(x)} = -\infty \text{ and } \lim_{x\downarrow -1} \frac{a(x)}{g^2(x)} = +\infty,$$

then the prerequisites of Theorem 3.8 are satisfied. ◊

The common denominator of the precceding two examples is that the coefficients in the dynamics of ρ fulfill the prerequisites of Theorem 3.8, which includes bounded derivatives. We now want to give an example where g does not have bounded derivatives in -1 and 1. We consider so-called Jacobi processes, which are given by the solution of

$$d\rho_t = \kappa(\theta - \rho_t)dt + \alpha\sqrt{1-\rho_t^2}\hat{W}_t. \tag{3.27}$$

Jacobi processes might be of interest for modelling stochastic correlation, because their stationary and transitional densities are well known and can be obtained quite explicit, see for example Boortz [2008].

By exploiting the boundary theory at the beginning of Section 3.4 or by checking when Condition (3.10) holds one can easily show that for $\kappa, \alpha > 0$ and θ such that

$$\kappa \geq \frac{\alpha^2}{1 \pm \theta}, \tag{3.28}$$

the boundaries -1 and 1 of the process defined in (3.27) are unattainable. Hence, we have that Assumption (H1) is fulfilled. We want to apply Proposition 3.20 and therefore have to check the boundedness of $\tilde{a}$ and upper boundedness of $\tilde{g}$. Note that $\tilde{g}$ turns into

$$\tilde{g}_u = \frac{2\rho_u}{(1-\rho_u^2)}a(\rho_u) + 2a'(\rho_u) + \frac{\alpha^2}{1-\rho_u^2} = \frac{2\rho_u}{(1-\rho_u^2)}(\kappa(\theta - \rho_u)) - 2\kappa + \frac{\alpha^2}{1-\rho_u^2},$$

and $\tilde{a}_u = 0$. In order to ensure upper boundedness of $\tilde{g}_u$ it is sufficient to show the existence of an $\varepsilon > 0$ such that $2\rho_u(\kappa(\theta - \rho_u)) + \alpha^2 < 0$, for all $|\rho_u| > 1 - \varepsilon$, P-almost surely. This is guaranteed by choosing $\kappa, \alpha > 0$ and θ such that the constants $\rho_{(1)}$ and $\rho_{(2)}$ defined by

$$\rho_{(1),(2)} = \frac{\theta}{2} \pm \sqrt{\frac{\theta^2}{4} + \frac{\alpha^2}{2\kappa}},$$

fulfill

$$-1 < \rho_{(1)} \leq \rho_{(2)} < 1. \tag{3.29}$$

Note, for example, that for $\alpha = 1$ and $\theta = 0.9$ Condition (3.28) is satisfied by $\kappa = 10$ and that this choice of parameters also fulfills (3.29).

3.5.2. Modelling correlation with Ornstein-Uhlenbeck processes

In the previous section we assumed that the stochastic correlation process is described in terms of the SDE (3.2). The correlation dynamics need not to be modelled directly. Alternatively, one can use a continuous bijection $b : (-1, 1) \to \mathbb{R}$, and model at first place the transformed process $b(\rho_t)$ as an SDE. This has the advantage that $b(\rho_t)$ can be modelled as a diffusion on $\mathbb{R}$ with Lipschitz coefficients. The correlation may be modelled as a standard mean reverting process, for example an Ornstein-Uhlenbeck process, the dynamics of which can be calibrated via standard methods.

In this section we will discuss this alternative approach of modelling correlation. As a paradigma example we will choose as bijection $b(x) = \frac{x}{\sqrt{1-x^2}}$, and we will assume that

$U_t = b(\rho_t)$ is an Ornstein-Uhlenbeck process with dynamics

$$dU_t = a(\theta - U_t)dt + \sigma_U d(\gamma dW_t^1 + \delta dW_t^2 + \sqrt{1-\gamma^2-\delta^2}dW_t^3),$$

where $a > 0$, $\theta \in \mathbb{R}$, $\sigma_U > 0$ and γ, $\delta \in (-1,1)$ are such that $\gamma^2 + \delta^2 \leq 1$. Notice that $\rho_t = \frac{U_t}{\sqrt{1+U_t^2}}$. We will prove that the prerequisites of Theorem 3.4 are satisfied and hence that the local risk minimization strategy is defined as in (3.9).

Lemma 3.25. *The correlation process ρ_t satisfies Conditions (H1) and (H2) and hence Theorem 3.4 holds in this setting.*

Proof. The proof is a simple application of Ito's formula. The first and the second derivative of $b^{-1} : \mathbb{R} \to]-1,1[, x \mapsto \frac{x}{\sqrt{1+x^2}}$ are given by $(b^{-1}(x))' = (1+x^2)^{-\frac{3}{2}}$ and $(b^{-1}(x))'' = -3x(1+x^2)^{-\frac{5}{2}}$. We set $\hat{W}_t = \gamma W_t^1 + \delta W_t^2 + \sqrt{1-\gamma^2-\delta^2}W_t^3$. We obtain

$$d\rho_t = (1-\rho_t^2)^{\frac{3}{2}}\sigma_U d\hat{W}_t + (1-\rho_t^2)\left(a\theta(1-\rho_t^2)^{\frac{1}{2}} - a\rho_t - \frac{3}{2}\rho_t(1-\rho_t^2)\sigma_U^2\right)dt.$$

It is straightforward to show that the coefficients of this SDE satisfy the conditions of Theorem 3.8. Hence the result. □

Bibliography

Carol Alexander. Optimal hedging using cointegration. *Philosophical Transactions of the Royal Society Series A*, 357:2039–2058, 1999.

Stefan Ankirchner and Gregor Heyne. Cross hedging with stochastic correlation. *Finance and Stochastics*, 16(1):17–43, 2012.

Stefan Ankirchner, Peter Imkeller, and Gonçalo Dos Reis. Classical and variational differentiability of BSDEs with quadratic growth. *Electronic Journal of Probability*, 12:no. 53, 1418–1453 (electronic), 2007.

Stefan Ankirchner, Peter Imkeller, and Gonzalo Dos Reis. Pricing and hedging of derivatives based on non-tradeable underlyings. *Mathematical Finance*, 20(2):289–312, 2010.

Stefan Ankirchner, Georgi Dimitroff, Gregor Heyne, and Christian Pigorsch. Futures cross-hedging with a stationary basis. *accepted for publication in Journal of Financial and Quantitative Analysis*, 2011.

Philippe Artzner, Freddy Delbaen, Jean Marc Eber, and Davidh Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

Louis Bachelier. Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure*, 3 (17):21–86, 1900.

Richard T. Baillie. Econometric tests of rationality and market efficiency. *Econometric Reviews*, 8:151–186, 1989.

Jocelyne Bion-Nadal. Time consistent dynamic risk processes. *Stochastic Processes and their Applications*, 119(2):633–654, 2009.

Jean-Michel Bismut. Conjugate convex functions in optimal stochastic control. *Journal of Mathematical Analysis and Applications*, 44:384–404, 1973.

Christopher K. Boortz. Modelling correlation risk. Master thesis. Technische Universität Berlin, 2008.

Robin J. Brenner and Kenneth F. Kroner. Arbitrage, cointegration, and testing the unbiasedness hypothesis in financial markets. *Journal of Financial and Quantitative Analysis*, 30:23–42, 1995.

Philippe Briand and Fulvia Confortola. BSDEs with stochastic Lipschitz condition and quadratic PDEs in Hilbert spaces. *Stochastic Processes and their Applications*, 118 (5):818–838, 2008.

Philippe Briand and Ying Hu. Quadratic BSDEs with convex generators and unbounded terminal conditions. *Probability Theory and Related Fields*, 141(3-4):543–567, 2008.

Luciano Campi and Walter Schachermayer. A Super-Replication Theorem in Kabanov's Model of Transaction Costs. *Finance and Stochastics*, 10(4):579–596, 2006.

Jaime Casassus and Pierre Collin-Dufresne. Stochastic convenience yield implied from commodity futures and interest rates. *Journal of Finance*, 60:2283–2331, 2005.

Kalok Chan. A further analysis of the lead-lag relationship between the cash market and stock index futures market. *Review of Financial Studies*, 5:123–152, 1987.

Patrick Cheridito and Mitja Stadje. Existence, minimality and approximation of solutions to BSDEs with convex drivers. *Stochastic Processes and their Applications*, 122 (4):1540–1565, 2012.

Patrick Cheridito, Freddy Delbaen, and Michael Kupper. Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, 11(3): 57–106, 2006.

Mark H. A. Davis. Optimal hedging with basis risk. In *From Stochastic Calculus to Mathematical Finance*, pages 169–187. Springer, Berlin, 2006.

Dimitry De Vallière, Emmanuel Denis, and Yuri Kabanov. Hedging of American Options under Transaction Costs. *Finance and Stochastics*, 13(1):105–119, 2009.

Freddy Delbaen. The structure of m-stable sets and in particular of the set of risk neutral measures. In *In Memoriam Paul-André Meyer - Seminaire de Probabilités XXXIX*, pages 215–258. Springer, 2006.

Freddy Delbaen and Walter Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300(3):463–520, 1994.

Freddy Delbaen and Walter Schachermayer. A compactness principle for bounded sequences of martingales with applications. In *Proceedings of the Seminar of Stochastic Analysis, Random Fields and Applications, Progress in Probability*, pages 137–173. Birkhauser, 1996.

Freddy Delbaen, Peter Grandits, Thorsten Rheinländer, Dominick Samperi, Martin Schweizer, and Christophe Stricker. Exponential hedging and entropic penalties. *Mathematical Finance*, 12(2):99–123, 2002.

Freddy Delbaen, Ying Hu, and Xiaobo Bao. Backward SDEs with superquadratic growth. *Probability Theory and Related Fields*, pages 1–48, 2010.

Claude Dellacherie and Paul-André Meyer. *Probabilities and Potential. B*, volume 72 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1982. ISBN 0-444-86526-8. Theory of martingales, Translated from the French by J. P. Wilson.

Samuel Drapeau, Gregor Heyne, and Michael Kupper. Minimal supersolutions of convex BSDEs. *preprint*, 2011.

Richard M. Dudley. Wiener functionals as Itô integrals. *Annals of Probability*, 5(1): 140–141, 1977.

Darrell Duffie and Larry G. Epstein. Stochastic differential utility. *Econometrica*, 60 (2):353–94, March 1992.

Darrell Duffie and Henry Richardson. Mean-variance hedging in continuous time. *Annals of Applied Probability*, 1:1–15, 1991.

Richard Durrett. *Probability: Theory and Examples.* Duxbury Press, Belmont, CA, second edition, 1996.

Nicole El Karoui and Marie-Claire Quenez. Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM Journal of Control and Optimization*, 33(1):29–66, 1995.

Nicole El Karoui, Christophe Kapoudjian, Etienne Pardoux, Shige Peng, and Marie-Claire Quenez. Reflected solutions of backward SDE's, and related obstacle problems for PDE's. *The Annals of Probability*, 25(2):702–737, 1997a.

Nicole El Karoui, Shige Peng, and Marie Claire Quenez. Backward stochastic differential equation in finance. *Mathematical Finance*, 1(1):1–71, January 1997b.

Robert F. Engle and Clive W. J. Granger. Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55:251–276, 1987.

Hans Föllmer and Irina Penner. Convex risk measures and the dynamics of their penalty functions. *Statistics & Decision*, 24(1):61–96, 2006.

Hans Föllmer and Alexander Schied. Convex measures of risk and trading constraint. *Finance and Stochastics*, 6(4):429–447, 2002.

Hans Föllmer and Alexander Schied. *Stochastic Finance. An Introduction in Discrete Time.* de Gruyter Studies in Mathematics. Walter de Gruyter, Berlin, New York, 2 edition, 2004.

Hans Föllmer and Dieter Sondermann. Hedging of nonredundant contingent claims. In *Contributions to Mathematical Economics*, pages 205–223. North-Holland, 1986.

Christoph Frei and Martin Schweizer. Exponential utility indifference valuation in two brownian settings with stochastic correlation. *Advances in Applied Probability*, 40(2): 401–423, 2008.

Clive W. J. Granger. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16:121–130, 1981.

J. Michael Harrison and Stanley R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3): 215–260, 1981.

Vicky Henderson. Valuation of claims on nontraded assets using utility maximization. *Mathematical Finance*, 12(4):351–373, 2002.

Gregor Heyne, Michael Kupper, and Christoph Mainberger. Minimal supersolutions of BSDEs with lower semicontinuous generators. *accepted for publication in Annales de l'Institut Henri Poincaré. Probabilités et Statistiques*, 2012.

Ulrich Horst, Traian A. Pirvu, and Gonçalo Dos Reis. On securitization, market completion and equilibrium risk transfer. *Mathematics and Financial Economics*, 2(4): 211–252, 2010.

Ying Hu, Peter Imkeller, and Matthias Müller. Utility maximization in incomplete markets. *Annals of Applied Probability*, 15 (3):1691–1712, 2005.

John C. Hull. *Options, Futures, and Other Derivatives.* Prentice Hall, Seventh edition, 2008.

Hardy Hulley and Thomas A. McWalter. Quadratic hedging of basis risk. Research Paper Series 225, Quantitative Finance Research Centre, University of Technology, Sydney, June 2008. URL `http://ideas.repec.org/p/uts/rpaper/225.html`.

Peter Imkeller, Anthony Réveillac, and Anja Richter. Differentiability of quadratic BSDEs generated by continuous martingales. *The Annals of Applied Probability*, 22 (1):285–336, 2012.

Olav Kallenberg. *Foundations of Modern Probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus (Graduate Texts in Mathematics).* Springer, August 2004.

Samuel Karlin and Howard M. Taylor. *A Second Course in Stochastic Processes.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1981.

Ira G. Kawaller, Paul D. Koch, and Timothy W. Koch. The temporal price relationship between S&P 500 futures and the S&P 500 index. *Journal of Finance*, 42:1309–1329, 1987.

Norihiko Kazamaki. *Continuous Exponential Martingales and BMO*, volume 1579 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 1994.

Magdalena Kobylanski. Backward stochastic differential equations and partial differential equations with quadratic growth. *Annals of Probability*, 28(2):558–602, 2000.

Michael Kupper and Walter Schachermayer. Representation Results for Law Invariant Time Consistent Functions. *Mathematics and Financial Economics*, 2(3):189–210, 2009.

Donald Lien and Xiangdong Luo. Estimating multiperiod hedge ratios in cointegrated markets. *Journal of Futures Markets*, 13:909–920, 1993.

Fabio Maccheroni, Massimo Marinacci, and Aldo Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6): 1447–1498, November 2006.

Michael Mania and Martin Schweizer. Dynamic exponential utility indifference valuation. *Annals of Applied Probability*, 15(3):2113–2143, 2005.

Michael Monoyios. Performance of utility-based strategies for hedging basis risk. *Quantitative Finance*, 4(3):245–255, 2004.

Marek Musiela and Thaleia Zariphopoulou. An example of indifference prices under exponential preferences. *Finance and Stochastics*, 8:229–239, 2004.

Victor K. Ng and Stephen Craig Pirrong. Price dynamics in refined petroleum spot and futures markets. *Journal of Empirical Finance*, 2:359–388, 1996.

David Nualart. *The Malliavin Calculus and Related Topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.

Etienne Pardoux and Shige Peng. Adapted solution of a backward stochastic differential equation. *System & Control Letters*, 14:55–61, 1990.

Shige Peng. Backward SDE and related g-expectation. *Backward Stochastic Differential Equation, Pitman Research Notes in Mathematics Series, 364, Longman, Harlow*, pages 141–159, 1997.

Shige Peng. Monotonic limit theorem of BSDE and nonlinear decomposition theorem of Doob–Meyer's type. *Probab. Theory Related Fields*, 113(4):473–499, 1999.

Philip E. Protter. *Stochastic Integration and Differential Equations*. Springer-Verlag, 2004.

Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

Ralph Tyrrel Rockafellar and Roger J-B Wets. *Variational Analysis*. Springer, Berlin, New York, 1998.

Ralph Tyrrell Rockafellar. Integral Functionals, Normal Integrands and Measurable Selections. In Jean Gossez, Enrique Lami Dozo, Jean Mawhin, and Lucien Waelbroeck, editors, *Nonlinear Operators and the Calculus of Variations*, volume 543 of *Lecture Notes in Mathematics*, pages 157–207. Springer Berlin / Heidelberg, 1976.

Eduardo S. Schwartz. The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, 52:923–73, 1997.

Martin Schweizer. Mean-variance hedging for general claims. *Annals of Applied Probability*, 2:171–179, 1992.

Martin Schweizer. A guided tour through quadratic hedging approaches. In *Option pricing, interest rates and risk management*, Handb. Math. Finance, pages 538–574. Cambridge Univ. Press, Cambridge, 2001.

Martin Schweizer. Local risk-minimization for multidimensional assets and payment streams. In *Advances in Mathematics of Finance*, pages 213–229. Banach Center Publications (83), 2008.

H. Mete Soner and Nizar Touzi. Stochastic target problems, dynamic programming, and viscosity solutions. *SIAM Journal on Control and Optimization*, 41(2):404–424 (electronic), 2002.

Hans R. Stoll and Robert E. Whaley. The dynamics of stock index and stock index futures returns. *Journal of Financial and Quantitative Analysis*, 25:441–468, 1990.

Cathrin van Emmerich. Modelling correlation as a stochastic process. Technical report, Bergische Universität Wuppertal, March 2006. URL `http://www-num.math.uni-wuppertal.de/fileadmin/mathe/www-num/preprints/amna_06_03.pdf`.

List of Figures

List of Tables